

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/38613 A2

(51) International Patent Classification⁷: C07K 19/00,
14/005, C12N 5/10, 15/33

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA,
ZW.

(21) International Application Number: PCT/EP01/12975

(22) International Filing Date:
9 November 2001 (09.11.2001)

(25) Filing Language: English

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

(30) Priority Data:
00124629.7 10 November 2000 (10.11.2000) EP
01109543.7 17 April 2001 (17.04.2001) EP
60/311,876 13 August 2001 (13.08.2001) US

Published:

- without international search report and to be republished upon receipt of that report
- entirely in electronic form (except for this front page) and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(71) Applicant: ARTEMIS PHARMACEUTICALS GMBH
[DE/DE]; Neurather Ring 1, 51063 Cologne (DE).

(72) Inventors: KÜHN, Ralf; Hellenthaler Str. 22, 50937
Köln (DE). FELDER, Susanne; Kopenhagener Str. 1,
41469 Neuss (DE). SCHWENK, Frieder; Konradenweg
15, 50999 Köln (DE). KÜTER-LUKS, Birgit; Niederich-
str. 10, 50668 Köln (DE). FAUST, Nicole; Dammelsfurter
Weg 9, 51063 Rösrath (DE).

(74) Agents: HELBING, Jörg et al.; Von Kreisler Selting
Werner, Postfach 10 22 41, 50462 Köln (DE).

WO 02/38613 A2

(54) Title: MODIFIED RECOMBINASE

(57) Abstract: The present invention concerns a fusion protein comprising a recombinase protein, preferably the site-specific DNA recombinase C31-Int of phage (C31, and a peptide sequence which directs the nuclear uptake of the fusion protein in eucaryotic cells, and the use of this fusion protein to recombine, invert or delete DNA molecules containing recognition sequences for said recombinase in eucaryotic cells at high efficiency. In addition the invention relates to a cell, preferably a mammalian cell which contains recognition sequences for said recombinase in its genome and wherein the genome is recombined by the action of said fusion protein. Moreover, the invention relates to the use of said cell to study the function of genes and for the creation of transgenic organisms to study gene function at various developmental stages, including the adult. In conclusion, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination.

Modified Recombinase

5 The present invention concerns a fusion protein comprising a recombinase protein, preferably the site-specific DNA recombinase C31-Int of phage Φ C31, and a peptide sequence which directs the nuclear uptake of the fusion protein in eucaryotic cells, and the use of this fusion protein to recombine, invert or delete DNA molecules containing recognition sequences for said recombinase in 10 eucaryotic cells at high efficiency. In addition the invention relates to a cell, preferably a mammalian cell which contains recognition sequences for said recombinase in its genome and wherein the genome is recombined by the action of said fusion protein. Moreover, the invention relates to the use of said cell to study the function of genes and for the creation of transgenic organisms to study 15 gene function at various developmental stages, including the adult. In conclusion, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination.

20 **Background of the invention**

The controlled and permanent modification of the genome of eucaryotic cells and organisms is an important method for research applications, e.g. for studying gene function, for medical applications like gene therapy and the creation of disease models and for the design of economically important animals and crops. 25 The basic methods for genome manipulations by the engineering of endogenous genes through gene targeting in murine embryonic stem (ES) cells are well established and used since many years (Capechi, Trends in Genetics, 5, 70-76 (1989)). Since ES cells can pass mutations induced *in vitro* to transgenic offspring *in vivo* it is possible to analyse the consequences of gene disruption in 30 the context of the entire organism. Thus, numerous mouse strains with functionally inactivated genes ("knock-out mice") have been created by this technology and utilised to study the biological function of a variety of genes (Koller et al., Ann. Rev. Immunol., 10, 705 - 730 (1992)). More importantly, mouse mutants created by this procedure (also known as "conventional, 35 complete or classical mutants"), contain the inactivated gene in all cells and

tissues throughout life. Thus, classical mouse mutants represent the best animal model for inherited human diseases as the mutation is introduced into the germline but are not the optimal model to study gene function in adults, e.g. to validate potential drug target genes.

5 A refined method of targeted mutagenesis, referred to as conditional mutagenesis, employs the Cre/loxP site-specific recombination system which enables the temporally and/or spatially restricted inactivation of target genes in cells or mice (Rajewsky et al., J. Clin. Invest., 98, 600 - 603 (1996)). The phage P1 derived Cre recombinase recognises a 34 bp sequence referred to as loxP site

10 which is structured as an inverted repeat of 13 bp separated by an asymmetric 8 bp sequence which defines the direction of the loxP site. If two loxP sites are located on a DNA molecule in the same orientation the intervening DNA sequence is excised by Cre recombinase from the parental molecule as a closed circle leaving one loxP site on each of the reaction products (Kilby et al., TIG, 9, 413-421 (1993)). The creation of conditional mouse mutants initially requires the generation of two mouse strains, one containing two or more Cre recombinase recognition (loxP) sites in its genome while the other harbours a Cre transgene. The former strain is generated by homologous recombination in ES cells as described above, except that the exon(s) of the target gene is (are) flanked by

20 two loxP sites which reside in introns and do not interfere with gene expression. The Cre transgenic strain contains a transgene whose expression is either constitutively active in certain cells and tissues or is inducible by external agents, depending on the promoter region used. Crossing of the loxP-flanked mouse strain with the Cre recombinase expressing strain enables the deletion of the

25 loxP-flanked exons in the genome of doubly transgenic offspring in a prespecified temporally and/or spatially restricted manner. Thus, the method allows the analysis of gene function in particular cell types and tissues of otherwise widely expressed genes. Moreover, it enables the analysis of gene function in the adult organism by circumventing embryonic lethality which is often the consequence of

30 complete (germline) gene inactivation. For pharmaceutical research, aiming to validate the utility of genes and their products for drug development, gene inactivation which is inducible in adults provides an excellent genetic tool as this mimicks the biological effects of target inhibition upon drug application.

Since the first description of the concept of conditional gene targeting using the Cre/loxP system in mice in 1994 (Gu et al., *Science* 265, 103-106 (1994)) this method became increasingly popular among the research community and resulted in a broad collection of genetic tools for biological research in the mouse.

5 More than 30 Cre transgenic mouse strains with various tissue specificities for gene inactivation have been created, including several "deleter" strains which allow to remove the loxP-flanked target gene segment in the male or female germline (Cohen-Tannoudji et al., *Mol. Hum. Reprod.* 4, 929-938 (1998); Metzger et al., *Curr. Op. Biotech.*, 10, 470-476 (1999)). The need to characterise
10 the expression pattern of Cre mediated recombination in newly generated strains stimulated the construction of a number of "Cre-reporter" strains which harbour a silent reporter gene the expression of which is activated upon Cre-mediated deletion (Nagy, *Genesis*, 26, 99-109 (2000)). Conditional mouse mutants have been reported for about 20 different genes, many of them could not be studied in
15 adults as their complete inactivation leads to embryonic lethality (Cohen-Tannoudji et al., *Mol. Hum. Reprod.* 4, 929-938 (1998)).

Great efforts have also been made to control the expression of Cre recombinase in an inducible fashion in mice. After the first demonstration that inducible gene
20 knock-out is feasible in adult mice using an interferon controlled promoter (Kühn et al., *Science*, 269, 1427-1429 (1995)), mainly two methods were applied to control the activity of Cre recombinase. First, it has been demonstrated that the fusion of Cre with the ligand binding domain of a mutant estrogen receptor allows to control recombinase activity by a specific steroid-like inducer. Several
25 transgenic mouse strains expressing such a fusion protein have been generated and allow to induce gene inactivation in specific tissues (Metzger et al., *Curr. Op. Biotech.*, 10, 470-476 (1999)). Furthermore, the tetracycline-regulated gene expression system has been successfully used to control the expression of Cre in transgenic mice and thus provides a second system for inducible gene
30 inactivation using doxycycline as inducer (Saam et al., *J. Biol. Chem.* 274, 38071-38082 (1999)).

In addition to the application of Cre/loxP for gene inactivation by deletion of a gene segment this recombination system has been proved to be useful also for a
35 number of other genomic manipulations in ES cells or mice. These include the

conditional activation of transgenes in mice, chromosome engineering to obtain deletion, translocation or inversion, the simple removal of selection marker genes, gene replacement, the targeted insertion of transgenes and the (in)activation of genes by inversion (Nagy, *Genesis*, 26, 99-109 (2000); Cohen-

5 Tannoudji et al., *Mol. Hum. Reprod.* 4, 929-938 (1998)). In conclusion, the Cre/loxP recombination system has been proven to be extremely useful for the analysis of gene function in mice by broadening the methodological spectrum for genome engineering. It can be expected that many of the protocols now established for the mouse may be applied in future also to other animals or

10 plants.

In contrast to the huge diversity of genome manipulations which have been developed for the Cre/loxP system, very limited efforts have been made to develop further site-specific recombination systems for the use in mammalian 15 cells. Alternative recombination systems of different specificity but with an efficiency comparable to Cre/loxP could further enhance the flexibility of genome engineering by the side to side use of two or more systems in the same cell or organism. Furthermore, unidirectional recombination systems which follow a different mechanism than the reversible Cre/loxP-mediated recombination should 20 allow to develop new applications for genome engineering which cannot be performed with the current systems.

The reasons for the almost exclusive use of the Cre/loxP system for site-specific 25 recombination in mammalian cells are readily explained by a number of requirements which must be fulfilled for the efficient use of a recombinase in mammalian cells:

- i) the recombinase should act independent of cofactors like helper proteins,
- ii) it should act independent of the supercoiling status of the target DNA 30 and also on mammalian chromatin,
- iii) it should be efficiently active and stable at a temperature of 37°C, and
- iv) it should recognize a target sequence which is sufficiently long to be unique among large genomes, and it should exhibit a very high affinity 35 to its target site for efficient action (Kilby et al., *TIG*, 9, 413-421 (1993)).

Among the more than 200 described members of the integrase and resolvase/invertase recombinase families only the Cre/loxP system is presently known to fulfill all of these requirements (Nunes-Düby et al., Nucleic Acids Res., 5 26, 391-406 (1998); Kilby et al., TIG, 9, 413-421 (1993); Ringrose et al., J. Mol. Biol., 284, 363 – 384 (1998)). Besides Cre/loxP a few recombinases have been shown to exhibit some activity in mammalian cells but their practical value is presently unclear as their efficiency has not been compared to the Cre/loxP system on the same genomic recombination substrate and in some cases it is 10 known that one or more of the criteria listed above are not met. The best characterised examples are the yeast derived FLP and Kw recombinases which exhibit a temperature optimum at 30°C but which are unstable at 37°C (Buchholz et al., Nature Biotech., 16, 657 – 662 (1998); Ringrose et al., Eur. J. Biochem., 248, 903 – 912). For FLP it has been shown in addition that its affinity to the FRT 15 target site is much lower as compared to the affinity of Cre to loxP sites (Ringrose et al., J. Mol. Biol., 284, 363 – 384 (1998)). Other recombinases which show in principle some activity in mammalian cells are a mutant integrase of phage λ , the integrases of phages Φ C31 and HK022, mutant $\gamma\delta$ -resolvase and β -recombinase (Lorbach et al., J. Mol. Biol., 296, 1175 – 81 (2000); Groth et al., 20 Proc. Natl. Acad. Sci. USA, 97, 5995 – 6000 (2000); Kolot et al., Mol. Biol. Rep. 26, 207 – 213 (1999); Schwikardi et al., FEBS Lett., 471, 147 – 150 (2000); Diaz et al., J. Biol. Chem., 274, 6634 – 6640 (1999)). Other phage integrase systems include coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase, CisA recombinase, XisF recombinase and transposon Tn4451 25 TnpX recombinase (Stark et al. Trends in Genetics 8, 432-439 (1992); Hatfull & Gridley, in Genetic Recombination. Eds. Kucherlapati & Smith, Am. Soc. Microbiol., Washington DC, 357-396 (1988)).

However, the practical value of these recombinases and integrases for use in 30 mammalian cells is limited as their efficiency to recombine mammalian genomic DNA has not been tested or compared with the Cre/loxP system. From the data available it can be assumed that these recombinases are much less effective than the Cre/loxP system.

In a few cases attempts have been made to improve the performance of recombinases in mammalian cells: for FLP a mutant showing improved thermostability and acticity at 37°C has been isolated but this mutant is still considerably more heat labile as compared to Cre (Buchholz et al., *Nature Biotech.*, 16, 657 – 662 (1998)). In the case of λ -integrase and $\gamma\delta$ -resolvase the absolute requirement for coproteins and supercoiled DNA could be eliminated by the introduction of specific point mutations (Schwikardi et al. *FEBS Lett* 471, pp147-50 (2000)).

10 The import of cytoplasmic proteins into the nucleus of eucaryotic cells through nuclear pores is a regulated, energy dependent process mediated by specific receptors (Görlich et al., *Science*, 271, 1513 – 1518 (1996)). Proteins which do not posses a signal sequence recognised by the nuclear import machinery are excluded from the nucleus and remain in the cytoplasm. Numerous of such 15 nuclear localisation signal sequences (NLS), which share a high proportion of basic amino acids in common, have been characterised (Boulikas, *Crit. Rev. Eucar. Gene Expression*, 3, 193 – 227 (1993)), the prototype of which is the 7 amino acid NLS derived from the T-antigen of the SV40 virus (Kalderon et. al, *Cell*, 39, 499 – 509 (1984)).

20 It was believed that the fusion of such an NLS peptide to a recombinase possibly would enhance the efficiency of the recombinase by mediating its import into the nucleus and therewith increasing the concentration of the recombinase inside the nucleus. However, for Cre recombinase it has been shown that the addition of 25 the SV-40 T-antigen NLS does not improve its recombination efficiency in mammalian cells (Le et al., *Nucleic Acid Res.*, 27, 4703 –4709 (1999)). Nevertheless, both Cre and a Cre-NLS-fusion protein are widely used. Schwikardi (Schwikardi et al., *FEBS Lett.* 471, pp147-50 (2000)) reported a $\gamma\delta$ -resolvase-SV-40 T-antigen NLS fusion protein, which also did not enhance the recombination 30 efficiency.

The level of activity exhibited by recombinases of diverse prokaryotic origin in mammalian cells may be the result of the intrinsic properties of an enzyme depending on parameters like its temperature optimum, its target site affinity, 35 protein structure and stability, the degree of cooperativity, the stability of the

synaptic complex and the dependence on coproteins or supercoiled DNA. Within the specific environment of mammalian cells the activity of a prokaryotic recombinase could be limited by additional factors such as a short half-life of the recombinase transcript, a short half-life of its protein, its inability to act on histone-complexed and higher order structured mammalian genomic DNA, exclusion from the nucleus or the recognition of cryptic splice sites within its mRNA resulting in a nonfunctional transcript. Due to the lack of information on the parameters listed above for almost all recombinases it is presently not possible to rationally optimise their performance in mammalian cells.

10

Summary of the Invention

The object to be solved by the invention of the present application is the provision of a recombination system alternative to the Cre/loxP system, which has a different specificity but an efficiency comparable to Cre/loxP. Such an alternative recombination system is particularly desirable for all those applications which require more than one potent recombination system for being successfully carried out (e.g. the methods disclosed in PCT/EP01/00060 and PCT/EP00/10162). Most surprisingly, it was found that the above object can be solved by fusing a signal peptide capable directing the nuclear import (hereinafter shortly referred to as nuclear localisation signal sequences (NLS)) to specific recombinases.

In contrast to the wildtype recombinases, the resulting modified recombinases allow a highly efficient recombination of extrachromosomal and chromosomal DNA in mammalian cells, and a highly efficient excision of extrachromosomal and chromosomal DNA-stretches, which are flanked by suitable recognition sites for said modified recombinases.

30 The present invention thus provides:

- (1) A fusion protein (hereinafter also referred to as "modified recombinase") comprising
 - (a) a recombinase domain comprising a recombinase protein or fragment thereof and
 - (b) a signal peptide domain being linked to (a) and directing the nuclear import

of said fusion protein in eucaryotic cells,

preferably the activity of the fusion protein in eucaryotic cells is significantly higher as compared to the activity of the wildtype recombinase corresponding to the recombinase of the recombinase domain;

- 5 (2) in a preferred embodiment of the fusion protein defined in (1) above, the recombinase domain comprises an integrase protein, preferably a phage Φ C31 integrase (C31-Int) protein or a mutant thereof;
- (3) a DNA coding for the fusion protein as defined in (1) or (2) above;
- (4) a vector containing the DNA as defined in (3) above;
- 10 (5) a microorganism containing the DNA of (3) above and/or the vector of (4) above;
- (6) a process for preparing the fusion protein as defined in (1) or (2) above which comprises culturing a microorganism as defined in (5) above;
- 15 (7) the use of the fusion protein as defined in (1) or (2) above to recombine DNA molecules, which contain recombinase recognition sequences for the recombinase protein of the recombinase domain, in eucaryotic cells;
- (8) a cell, preferably a mammalian cell containing the DNA sequence of (3) above in its genome;
- 20 (9) the use of the cell of (8) above for studying the function of genes and for the creation of transgenic organisms;
- (10) a transgenic organism, preferably a transgenic mammal containing the DNA sequence of (3) above in its genome;
- (11) the use of the transgenic organism of (10) above for studying gene function at various developmental stages; and
- 25 (12) a method for recombining DNA molecules of cells or organisms containing recognition sequences for the recombinase protein of the recombinase domain as defined in (1) or (2) above, which method comprises supplying the cells or organisms with a fusion protein as defined in (1) or (2) above, or with a DNA sequence of (3) above and/or a vector of (4) above which are capable of
- 30 expressing said fusion protein in the cell or organism.

The present invention combines the use of prokaryotic recombinases such as the C31-Int with a eukaryotic signal sequence which increases its efficiency in mammalian cells such that it is equal to the widely used Cre/loxP recombination system. The improved recombination system of the present invention provides

an alternative recombination system for use in mammalian cells and organisms which allows to perform the same types of genomic modifications as shown for Cre/loxP, including conditional gene inactivation by recombinase-mediated deletion, the conditional activation of transgenes in mice, chromosome 5 engineering to obtain deletion, translocation or inversion, the simple removal of selection marker genes, gene replacement, the targeted insertion of transgenes and the (in)activation of genes by inversion.

Short Description of Figures

10 Fig. 1: C31-Int and Cre recombinase expression vectors and a recombinase reporter vector used for transient and stable transfections

Fig. 2: Results of transient transfections of C31 Int and Cre expression vectors and reporter vectors into CHO cells.

15 Fig. 3: Results of transient transfections of XisA and Ssv recombinase expression vectors with and without nuclear localisation signals and reporter vectors into CHO cells.

20 Fig. 4: Results of transient transfections of C31-Int and Cre recombinase vectors into a stable reporter cell line.

Fig. 5: In situ detection of β -galactosidase in 3T3(pRK64)-3 cells transfected with recombinase expression vectors

25 Fig. 6: Test vector for C31-Int mediated deletion, pRK64, and the expected deletion product.

30 Fig. 7: PCR products generated with the primers P64-1 and P64-4 and sequence comparison.

Fig. 8: ROSA26 locus of the C31 reporter mice carrying a C31 reporter construct.

35 Fig. 9: In situ detection of β -galactosidase in a cryosection of the testis of: (A) a double transgenic mouse carrying both the recombinase and the reporter; and

(B) a transgenic mouse carrying only the reporter as a control.

Detailed Description of the Invention

The "organisms" according to the present invention are multi-cell organisms and can be vertebrates such as mammals (humans and non-human animals including rodents such as mice or rats) or non-mammals (e.g. fish), or can be invertebrates such as insects or worms, or can be plants (higher plants, algi or fungi). Most preferred living organisms are mice and fish.

5 "Cells" and "eucaryotic cells" according to the present invention include cells isolated from the above defined living organism and cultured *in vitro*. These cells can be transformed (immortalized) or untransformed (directly derived from the living organism; primary cell culture).

10 "Microorganism" according to the present invention relates to procaryotes (e.g. *E. coli*) and eucaryotic microorganisms (e.g. yeasts).

15 According to embodiment (1) of the present invention, the activity of the fusion protein in eucaryotic cells is significantly higher as compared to the acitivity of the wildtype recombinase corresponding to the recombinase of the recombinase domain. A "significantly higher activity" in accordance with the present invention refers to an increase in activity of at least 50%, preferably at least 75%, more preferably at least 100% relative to the corresponding wildtyp recombinase in eucaryotic cells. A "significantly higher activty" also implies that the resulting 20 fusion protein has at least 25%, preferably at least 50% and more preferably at least 75%, of the activity of Cre/loxP in 3T3 cells with a stably integrated target sequence.

25 Recombinase proteins which can be used in the recombinase domain of the fusion protein of the present invention (i.e., giving a fusion having a "significantly higher activty" as defined above) include , but are not limited to, a certain type of recombinases belonging to the family of of large serine recombinases (Thorpe et al., Control of directionality in the site-specific recombination system of the streptomyces phage φC31, Molecular Microbiology 38(2), 232-241 (2000)). This 30 family includes bacteriophage φC31 integrase ("C31-Int"; the amino acid

sequence of said integrase and a DNA sequence coding therefor are shown in SEQ ID NOs:21 and 20, respectively), coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase ("R4 Sre" deposited under GI 793758; the amino acid sequence of said recombinase and a DNA sequence 5 coding therefor are shown in SEQ ID NOs:55 and 54, respectively), bacillus subtilis CisA recombinase ("CisA" deposited under GI 142689; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:57 and 56, respectively), XisF recombinase from annabaena sp. Strain PCC 7120 (Cyanobacterium; "XisF" deposited under GI 349678; the amino 10 acid sequence of said integrase and a DNA sequence coding therefor are shown in SEQ ID NOs:59 and 58, respectively), transposon Tn4451 TnpX recombinase ("TnpX" deposited under GI 551135; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:61 and 60, respectively), "XisA" recombinase from annabaena sp. Strain PCC 7120 15 (Cyanobacterium; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:63 and 62, respectively), "SSV" recombinase from phage of sulfolobus shibatae (the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:65 and 64, respectively); lactococcal bacteriophage TP901-1 recombinase 20 (TP901-1 complete genome deposited under GI 13786531; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:108 and 107, respectively), and the like, or mutants thereof. Other procaryotic recombinases known in the art are also applicable.

25 A "mutant" of the above recombinases in accordance with the present invention relates to a mutant of the respective original (viz. wild-type) recombinase having a recombinase activity similar (e.g. at least about 90%) to that of said wild-type recombinase. Mutants include truncated forms of the recombinase (such as N- or C-terminal truncated recombinase proteins), deletion-type mutants (where one 30 or more amino acid residues or segments having more than one continuous amino acid residue have been deleted from the primary sequence of the wildtyp recombinase), replacement-type mutants (where one or more amino acid residues or segments of the primary sequence of the wildtyp recombinase have been replaced with alternative amino acid residues or segments), or 35 combinations thereof.

According to embodiment (2) of the invention, the recombinase domain comprises an integrase protein, preferably a phage Φ C31 integrase (C31-Int) protein or a mutant thereof. Thus, the present invention provides a fusion protein

5 comprising

(a) an integrase domain being a C31-Int protein or a mutant thereof, and
(b) a signal peptide domain being linked to (a) and directing the nuclear import of said fusion protein into eucaryotic cells.

10 In the fusion protein of embodiment (2), the integrase domain is preferably a C31-Int having the amino acid sequence shown in SEQ ID NO:21 or a C-terminal truncated form thereof. Suitable truncated forms of the C31-Int comprise amino acid residues 306 to 613 of SEQ ID NO:21.

15 The signal peptide domain (hereinafter also referred to as "NLS") is preferably derived from yeast GAL4, SKI3, L29 or histone H2B proteins, polyoma virus large T protein, VP1 or VP2 capsid protein, SV40 VP1 or VP2 capsid protein, Adenovirus E1a or DBP protein, influenza virus NS1 protein, hepatitis virus core antigen or the mammalian lamin, c-myc, max, c-myb, p53, c-erbA, jun, Tax, steroid
20 receptor or Mx proteins (see Boulikas, Crit. Rev. Eucar. Gene Expression, 3, 193 - 227 (1993)); simian virus 40 ("SV40") T-antigen (Kalderon et. al, Cell, 39, 499 - 509 (1984)) or other proteins with known nuclear localisation. The NLS is preferably derived from the SV40 T-antigen.

25 Furthermore, the signal peptide domain preferably has a length of 5 to 74, preferably 7 to 15 amino acid residues. More preferred is that the signal peptide domain comprises a segment of 6 amino acid residues wherein at least 2 amino acid residues, preferably at least 3 amino acid residues are positively charged basic amino acids. Basic amino acids include, but are not limited to, Lysin, Arginin
30 and Histidine. Particularly preferred signal peptides are show in the following table.

Organism	Sequence/(SEQ ID NO:)
yeast GAL4	MKx11CRLKKLKCSKEPKCAKCLKx5Rx3KTKR (24)
35 yeast SKI3	IKYFKKKFPKD (25)

	13	
yeast L29		MTGSKTRKHRGSGA (26)
yeast histone H2B		(MTGSKHRKHPGSGA) (27)
polyoma virus large T protein		(GKKRSKA) (28)
5 polyoma virus VP1 capsid protein		(PKKAREDVSRKPR) (29)
polyoma virus VP2 capsid protein		(APKRKSGVSKC) (30)
SV40 VP1 capsid protein		(EEDGPQKKKRRL) (31)
SV40 VP2 capsid protein,		(APTKRKGS) (32)
Adenovirus E1a protein		(PNKKKRK) (33)
10		(KRPRP) (34)
Adenovirus DBP protein		(CGGLSSKRPRP) (35)
Influenza virus NS1 protein		(PPKKRMRRRIEPKKKKKRP) (36)
		(PFLDRLRRDQK) (37)
human laminA		(PKQKRKMAR) (38)
15 human c-myc		(SVTKKRKLE) (39)
		(CGGAAKRVKLD) (40)
		(PAAKRVKLD) (41)
		(RQRRNELKRSP) (42)
HUMAN max		(PQSRKKLR) (43)
HUMAN c-myb		(PLKKKIKQ) (44)
20 HUMAN p53		(PQPKKKP) (45)
HUMAN c-erbA		(SKRVAKRKL) (46)
VIRAL jun		(ASKSRKRKL) (47)
HUMAN Tax		(GGLCSARLHRHALLAT) (48)
Mammalian glucocorticoid receptor		(RKTKKKIK) (49)
25 HUMAN ANDROGEN RECEPTOR		(RKLKKLGN) (50)
MAMMALIAN ESTROGEN RECEPTOR		(RKDRRGGR) (51)
Mx proteins		(DTREKKKFLKRRLLRLDE) (52)
SV40 T-antigen		(PKKKRKV) (53)
30		The most preferred signal peptide domain is that of SV40 T-antigen having the sequence Pro-Lys-Lys-Lys-Arg-Lys-Val.

The signal peptide domain may be linked to the N-terminal or C-terminal of the integrase domain or may be integrated into the integrase domain, preferably the signal peptide domain is linked to the C-terminal of the integrase domain. With

regard to phage Φ C31 integrase protein of embodiment (2) of the invention it was found that the fusion of an NLS-peptide to the C-terminus of the integrase provided a much higher increase of activity as compared to the fusion of the same NLS-peptide to the N-terminus of the integrase (see Example 1, figures 3 and 4).

According to the present invention, the signal peptide domain may be linked to the integrase domain directly or through a linker peptide. Suitable linkers include peptides having from 1 to 30, preferably 1 to 15 amino acid residues, said amino acid residues being essentially neutral amino acids such as Gly, Ala and Val.

The most preferred fusion protein of the present invention comprises the amino acid sequence shown in SEQ ID NO:23 (a suitable DNA sequence coding for said fusion protein being shown in SEQ ID NO:22).

Further preferred fusion proteins of the present invention are "NLS-XisA" and "NLS-SSV" (having the NLS-peptide fused to the N-terminus of the recombinases) as shown in SEQ ID NO:67 and 69, respectively (suitable DNA sequences coding for said fusion proteins being shown in SEQ ID NO:66 and 68, respectively).

In embodiments (7), (8), (10) and (12) of the invention the DNA molecules, the cell or transgenic organism may also contain recognition sequences for the recombinase protein of the recombinase domain. Thus, when utilizing the fusion protein of embodiment (2), the C31-Int recognition sequences attP and attB are present in DNA molecules, the cell or transgenic organism.

The term "mammal" as used in embodiment (10) of the invention includes non-human mammals (viz. animals as defined above) and humans (if such subject matter is patentable with the respective patent authority).

Since the modified recombinase of the invention, in particular the modified C31-Int, acts in mammalian cells as efficient (or at least almost as efficient) as the widely used Cre/loxP system it can be used for a large variety of genomic modifications (including the methods disclosed in PCT/EP01/00060 and

PCT/EP00/10162, the content of which is herewith incorporated by reference). Concerning embodiment (11) it is to be noted that the mammals of embodiment (10) can be used to study the function of genes, e.g. in mice, by conditional gene targeting. For this purpose suitable recognition sequences - when utilizing the 5 fusion protein of embodiment (2), one attP and one attB site (C31-Int recognition sequences) in the same orientation - can be introduced into introns of a gene by homologous recombination of a gene targeting vector in ES cells such that the two sites flank one or more exons of the gene to be studied but do not interfere with gene expression. A selection marker gene, needed to isolate recombinant ES 10 cell clones, can be flanked by two recognition sites of another recombinase such as loxP or FRT sites to enable deletion of the marker gene upon transient expression of the respective recombinase in ES cells. These ES cells can be used to generate germline chimaeric mice which transmit the target gene modified by att sites to their offspring and allow to establish a modified mouse strain. The 15 crossing of this strain with a C31-Int recombinase transgenic line or the application of C31-Int protein will result in the deletion of the att-flanked gene segment from the genome of doubly transgenic offspring and the inactivation of the target gene in doubly transgenic offspring in a prespecified temporally and/or spatially restricted manner. The C31-Int transgenic strain contains a transgene 20 whose expression is either constitutively active in certain cells and tissues or is inducible by external agents, depending on the promoter region used. If an attB and an attP site are placed into the genome in opposite orientation C31-Int mediated recombination results in the irreversible inversion of the flanked gene segment leading the functional loss of one or more exons of the target gene. 25 Thus, the method allows the analysis of gene function in particular cell types and tissues of otherwise widely expressed genes and circumvents embryonic lethality which is often the consequence of complete (germline) gene inactivation. For the validation of genes and their products for drug development, gene inactivation which is inducible in adults provides an excellent genetic tool as this mimicks the 30 biological effects of target inhibition upon drug application. If a pair of attB/P sites is placed in the same or opposite orientation into a chromosome at large distance using two gene targeting vectors, C31-Int recombination allows to delete or invert chromosome segments containing one or more genes, or chromosomal translocations if the two sites are located on different 35 chromosomes. In another application of the method a pair of attB/P sites is

placed in the same orientation within a transgene such that the deletion of the att-flanked DNA segment results in gene expression, e.g. of a toxin or reporter gene for cell lineage studies, or in the inactivation of the transgene.

5 In addition, according with embodiment (12) of the invention, the recombination system of embodiment (1), in particular the C31-Int recombination system of embodiment (2), can also be used for the site specific integration of foreign DNA into the genome of mammalian cells, e.g. for gene therapy. For this purpose, and if the C31-Int recombination system of embodiment (2) is utilized, only one attB
10 (or attP) site is initially introduced into the genome by homologous recombination, or an endogenous genomic sequence which resembles attB or attP is used. The application of a vector containing an attP (or attB) site to such cells or mice in conjunction with the expression of C31-Int recombinase will lead to the site specific integration of the vector into the genomic att site.

15

Thus, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination. Said process possesses the following advantages over current technology:

20 (i) the modified recombinase, in particular the modified C31-Integrase, allows to recombine extrachromosomal and genomic DNA in mammalian cells at much higher efficiency as compared to the use of its wildtype form;

25 (ii) the modified recombinase, in particular the modified C31-Integrase, is the first described alternative recombination system with equal efficiency to Cre/loxP for the recombination of chromosomal DNA in mammalian cells.

The appended figures further explain the present invention:

30 Figure 1 shows C31-Int and Cre recombinase expression vectors and a recombinase reporter vector used for transient and stable transfections.

A-D: Mammalian expression vectors for recombinases which contain the CMV immediate early promoter followed by a hybrid Intron, the coding region of the recombinase to be tested, and an artificial polyadenylation signal sequence (pA).

A: pCMV-C31Int(wt) containing the nonmodified (wildtype) 1.85 kb coding region of C31-Int as found in the genome of phage Φ X31.

B: pCMV-C31Int(NNLS) containing a modified C31-Int gene coding for the full length C31-Int protein with a N-terminal fusion to the SV40 virus large T antigen 5 nuclear localisation signal (NLS).

C: pCMV-C31Int(CNLS) containing a modified C31-Int gene coding for the full length C31-Int protein with a C-terminal fusion to the SV40 virus large T antigen nuclear localisation signal (NLS).

D: pCMV-Cre contains the 1.1 kb Cre coding region with an N-terminal fusion to 10 the SV40 T antigen NLS.

E: Recombination substrate vector pRK64 contains a SV40 promoter region followed by a 1.1 kb cassette consisting of the coding region of the puromycin 15 resistance gene and a polyadenylation signal sequence, flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site of C31-Int. pRK64 contains in addition two Cre recognition (loxP) sites in direct orientation next to the att sites.

Figure 2 shows results of transient transfections of C31-Int and Cre recombinase and reporter vectors into CHO cells.

All transfections were performed with a fixed amount of the reporter plasmid 20 pRK64 and 0.5 ng or 1 ng of the recombinase expression plasmids pCMV-C31-Int(wt) (samples 4-5), pCMV-C31-Int(NNLS) (samples 6-7), pCMV-C31-Int(CNLS) (samples 8-9) or pCMV-Cre (samples 10-11). Negative controls: transfection with pRK64 (sample 3) or pUC19 alone (sample 1). Positive control: transfection with the Cre-recombined reporter pRK64(Δ Cre) (sample 2).

25 The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)), the RLU from the assay for Luciferase, the ratio of the β -galactosidase and Luciferase values with standard deviation (RLU $\times 10^5$ (Gal/Luc)), and the relative activity of the various recombinases as compared to the positive control defined 30 as 1.

Figure 3 shows results of transient transfections of XisA and Ssv recombinases and reporter vectors into CHO cells.

All transfections were performed with fixed amounts of the reporter plasmids 35 pPGKnif (for XisA) and pPGKattA (for SSV) and 25 ng or 100 ng of the

recombinase expression plasmids pCMV-XisA, pCMV-XisA(NNLS) and 10 ng or 20 ng of the expression plasmids pCMV-Ssv and pCMV-Ssv(NNLS). Negative controls: transfection with pPGKnif or pPGKattA alone.

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)), the RLU from the assay for Luciferase, the ratio of the β -galactosidase and "Luciferase" values with standard deviation (RLU $\times 10^5$ (Gal/Luc)).

Figure 4 shows results of transient transfections of recombinase vectors into a stable reporter cell line.

All transfections were performed with a NIH 3T3 derived clone containing stably integrated copies of the pRK64 recombination substrate vector. Either 32 ng or 64 ng of the recombinase expression plasmids pCMV-C31-Int(wt) (samples 2-3), pCMV-C31-Int(NNLS) (samples 4-5), pCMV-C31-Int(CNLS) (samples 6-7) or pCMV-Cre(NNLS) (samples 8-9). Negative control: transfection with pUC19 alone (sample 1).

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)) and the relative activity of the various recombinases as compared to the value obtained with pCMV-Cre(NNLS) defined as 1.

Figure 5 shows the in situ detection of β -galactosidase in 3T3(pRK64)-3 cells transfected with recombinase expression vectors.

The Cre and C31-Int recombinase reporter cell line 3T3(pRK64)-3 was either not transfected with DNA (A), transfected with the Cre expression vector pCMV-Cre (B) or with the C31-Int expression vector pCMV-C31-Int(CNLS). Two days after transfection the cells were fixed and incubated with the histochemical X-Gal assay which develops a blue stain in β -galactosidase positive cells indicating recombinase mediated activation of the reporter gene.

Figure 6 shows the test vector for C31-Int mediated deletion, pRK64, and the expected product of deletion, pRK64(Δ Int).

Plasmid pRK64 contains the 1.1 kb cassette of the coding region of the puromycin resistance gene and a polyadenylation signal, which is flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site (large triangles) of C31-

Int. These attB and attP sites are oriented in the same way to each other (thick black arrows) which is used by the Φ X31 phage to integrate into the bacterial genome. In addition, the cassette is flanked by two Cre recombinase recognition (loxP) sites in the same orientation (black small triangles). For better orientation 5 the half sites of the att sequences are labelled by a direction (thin arrow) and numbered 1-4. The 3 bp sequence within the att sites at which recombination occurs is framed by a box. The positions at which the PCR primers P64-1 and P64-4 hybridise to the pRK64 vector are indicated by arrows, pointing into the 3' direction of both oligonucleotides.

10 PRK64(Δ Int) depicts the deletion product expected from the C31-Int mediated recombination between the att sites of pRK64. The recombination between a pair of attB/attP sites generates an attR site remaining on the parental DNA molecule while the puromycin cassette is excised. In this configuration the primers P64-1 and P64-4 will amplify a PCR product of 630 bp from pRK64(Δ Int).

15

Figure 7 shows PCR products generated with the primers P64-1 and P64-4 and a sequence comparison of the PCR product.

A: Analysis of PCR products on an agarose gel from PCR reactions using the Primers P64-1 and P64-4 on DNA extracted from MEF5-5 cells transfected 2 days 20 before with plasmid pRK64 alone (lane 4), with pRK64 + CMV-Cre (lane 3), with pRK64 + pCMV-C31-Int(wt) (lane 2), and from a control reaction which did not contain cellular DNA (lane 1). The product with an apparent size around 650 bp, as compared to the size marker used, from lane 2 was excised from the agarose gel and purified. The PCR product was cloned into a sequencing plasmid vector 25 and gave rise to the plasmid pRK80d. The insert of this plasmid was sequenced using reverse primer (seq80d) and compared to the predicted sequence of the pRK64 vector after C31-Int mediated deletion of the att flanked cassette, pRK64(Δ Int). The cloned PCR product shows a 100% identity with the predicted attR sequence after deletion. The generated attR site is shown in a box, with the 30 same sequence designation used in Figure 5. The nucleotide positions (pos.) of the compared sequences pRK64(Δ Int) and Seq80d are indicated.

Figure 8 shows the modified ROSA26 locus of C31 reporter mice (Seq ID NO:106). A recombination substrate has been inserted in the ROSA26 locus. The 35 substrate consists of a splice acceptor (SA) followed by a cassette consisting of

the hygromycin resistance gene driven by a PGK promoter and flanked by the recombination sites attB and attP. In addition the reporter contains two Cre recognition sites (loxP) in direct orientation next to the att sites. This cassette is followed by the coding region for β -galactosidase, which is only expressed when

5 the hygromycin resistance gene has been deleted by recombination.

Figure 9 shows the *in situ* detection of β -galactosidase activity. A cryosection of the testis of a double transgenic mouse carrying both the C31-int recombinase and the recombination substrate was stained with X-Gal (A). The blue colour

10 indicates recombination of the substrate, which leads to the expression of β -galactosidase. As a control a cryosection of testis of a transgenic mouse carrying only the recombination substrate was stained with X-Gal (B).

The present invention is further illustrated by the following Examples which are,

15 however, not to be construed as to limit the invention.

Examples

Example 1

20 As compared to Cre recombinase the wildtype form of C31-Int exhibits a significantly lower recombination activity in mammalian cells which falls in the range of 10 – 40% of Cre, depending on the assay system used (see below). As a measure which may increase C31-Int efficiency in eukaryotic cells we designed mammalian expression vectors for N- or C-terminal fusion proteins of C31-Int

25 with a peptide was designed which is recognised by the nuclear import machinery. The recombination efficiency obtained by this modified C31-Int recombinase in mammalian cells was compared side by side to the unmodified (wildtype) form of C31-Int and to Cre recombinase. For the quantification of recombinase activities the expression vectors were transiently introduced into a

30 mammalian cell line together with a reporter vector which contains C31-Int and Cre target sites and leads to the expression of β -galactosidase upon recombinase mediated deletion of a vector segment flanked by recombinase recognition sites.

A. Plasmid constructions:

Construction of the recombination test vectors pPGKnif and pPGKattA: first a nifD site (Haselkorn, Annu Rev.Genet. 26, 113-130 (1992)) generated by the annealing of the two synthetic oligonucleotides nifD3 (SEQ ID NO:89) and nifD4 (SEQ ID NO:90), was ligated into the BamHI restriction site of the vector PSV-
5 Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)), 3' of its puromycin resistance gene and loxP site, giving rise to plasmid pPGKnifD3' (SEQ ID NO:79). Next, another nifD site, generated by the annealing of the two synthetic oligonucleotides nifD1 (SEQ ID NO:87) and nifD2 (SEQ ID NO:88), was ligated into the BstBI restriction site of plasmid pPGKnifD3', upstream of the
10 puromycin resistance gene and loxP site, giving rise to plasmid pPGKnifD (SEQ ID NO:78). For pPGKattA (Muskhelishvili et al., Mol.Gen.Genet. 237, 334-342 (1993)) first a 352bp-fragment was amplified from genomic DNA from the thermophilic bacterium *Sulfolobus shibatae* (DSM-5389, DSMZ Braunschweig-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder
15 Weg 1b, D-38124 Braunschweig, Germany) with oligonucleotides SSV5 (SEQ ID NO:96) and SSV6 (SEQ ID NO:97) including restriction sites for BamHI and BstBI. The amplified fragment was cloned into the BamHI site of the vector PSV-Pax1 giving rise to plasmid pPGKattA1 (SEQ ID NO:82), subsequently the same 352 bp-fragment was cloned into the BstBI site of pPGKattA1 giving rise to the
20 plasmid pPGKattA2 (SEQ ID NO:83). The sequence and orientation of both nifD sites and attA sites was confirmed by DNA sequence analysis. In pPGKnifD/pPGKattA2 the newly cloned nifD/attA sites (positions 535-619 and 1722-1787/ positions 6718-7081 and 12-363) are in the same orientation flanking the puromycin resistance gene and the SV40 early polyadenylation
25 sequence. The nifD/attA sites are followed by loxP sites in the same orientation (positions 623 - 656 and 1794 - 1827/ positions 7085-7118 and 369-402). The puromycin cassette is transcribed from the SV40 early enhancer/promoter region and followed by the coding region for *E. coli* β -galactosidase and the SV40 late region polyadenylation sequence.

30 Construction of XisA and SSV expression vectors: First the XisA gene of cyanobacterium PCC7120 was amplified by PCR from genomic DNA from *Nostoc* strain PCC7120 (CNCM-Collection Nationale de Cultures de Microorganismes, Institut Pasteur, Paris) using the primers XisA1 (SEQ ID NO:84) and XisA3 (SEQ
35 ID NO:86), and XisA1 (SEQ ID NO:84) and XisA2 (SEQ ID NO:85) (with NLS).

The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmids pRK42a and pRK43 (with NNLS). The DNA sequence of the insert was determined and found to be identical to the published XisA sequence (Genbank GI:3953452) apart from four silent point mutations. The XisA gene was isolated as a 1.4 kb fragment from pRK42a and pRK43 by digestion with NotI and ligated into the generic mammalian expression vector pRK50 (see below), opened with NotI, giving rise to the XisA expression vectors pCMV-XisA (SEQ ID NO:76) and pCMV-XisA(NNLS) (SEQ ID NO:77). pCMV-XisA(wt) contains a Cytomegalovirus immediated early gene promoter (position 1 – 616), a 240 bp hybrid intron (position 716 – 953), the XisA gene (position 974 – 2392), and a synthetic polyadenylation sequence (position 2413 – 2591).

The SSV gene was amplified from genomic DNA from the thermophilic bacterium *Sulfolobus shibatae* (DSM-5389, DSMZ Braunschweig- Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) in two PCR steps because of an internal attP sequence. First, two overlapping PCR fragments were created with the oligonucleotides SSV1-1 (SEQ ID NO:91) (or SSV1-2 for the SSV(NNLS) gene) and SSV2 (SEQ ID NO:93) and oligonucleotides SSV3 (SEQ ID NO:94) and SSV4 (SEQ ID NO:95). Using these overlapping fragments as template, a 1000bp fragment containing the complete SSV coding sequence was amplified with primers SSV1-1 (or SSV1-2 for the SSV(NNLS) gene) and SSV4. The 5' 620 bp-fragments of these PCR products were isolated by digestion with NotI-XhoI and cloned into vector pBluescript II KS giving rise to plasmids pRK47 and pRK48 (with NLS). The 3' 380 bp fragment generated by XhoI-digestion was cloned into the XhoI restriction site of vector pBluescript II KS giving rise to the plasmid pBS-SSVs (SEQ ID NO:72). The 380bp SSV-fragment was then isolated by digestion of pBS-SSVs with XhoI and ligated into pRK47 and pRK48 opened by XhoI giving rise to plasmids pBS-SSV3 (SEQ ID NO:70) and pBS-SSV4 (SEQ ID NO:71) (with NLS) containing the complete SSV gene. Sequencing of the plasmids confirmed one point mutation in both plasmids. Therefore 312 bp/ 91 bp fragments generated by digestion with EcoRV-SmaI/ EcoRV-XhoI of another clone of pRK47 were exchanged in plasmids pBS-SSV3/ pBS-SSV4. Sequences were confirmed by sequencing. The SSV gene was isolated from pRK47 and pRK48 by digestion with NotI and KpnI and ligated into the generic mammalian expression vector

pRK50 (see below), opened with *NotI* and *SalI*, giving rise to the SSV expression vectors pCMV-SSV(wt) (SEQ ID NO:74) and pCMV-SSV(NNLS) (SEQ ID NO:75).

Construction of the recombination test vector pRK64: first an attB site (Thorpe et al. Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)), generated by the annealing of the two synthetic oligonucleotides C31-4 (SEQ ID NO:1) and C31-5 (SEQ ID NO:2), was ligated into the *BstBI* restriction site of the vector PSV-Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)), 5' of its puromycin resistance gene and loxP site, giving rise to plasmid pRK52. The sequence and orientation of the cloned attB site was confirmed by DNA sequence analysis. Next, an attP site (Thorpe et al. Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)), generated by the annealing of the two synthetic oligonucleotides C31-6 (SEQ ID NO:3) and C31-7-2 (SEQ ID NO:4), was ligated into the *BamHI* restriction site of plasmid pRK52, downstream of the puromycin resistance gene and loxP site, giving rise to plasmid pRK64 (SEQ ID NO:5). The sequence and orientation of the attP site was confirmed by DNA sequence analysis. In pRK64 the newly cloned attB (position 348 – 431) and attP (position 1534 – 1617) sites are in the same orientation flanking the puromycin resistance gene and the SV40 early polyadenylation sequence. The attB and attP sites are followed by loxP sites in the same orientation (positions 435 – 469 and 1624 – 1658). The puromycin cassette is transcribed from the SV40 early enhancer/promoter region and followed by the coding region for *E. coli* β -galactosidase and the SV40 late region polyadenylation sequence.

Construction of C31-Int expression vectors: First the C31-Int gene of phage ϕ C31 was amplified by PCR from phage DNA (DSM-49156, DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) using the primers C31-1 (SEQ ID NO:6) and C31-3 (SEQ ID NO:7). The ends of the PCR product were digested with *NotI* and the product was ligated into plasmid pBluescript II KS, opened with *NotI*, giving rise to plasmid pRK40. The DNA sequence of the 1.85 kb insert was determined and found to be identical to the published C31-Int gene (Kuhstoss et al., J. Mol. Biol. 222, 897-908 (1991)), except for an error in the stop codon. This error was repaired by PCR amplification of a 300 bp fragment from plasmid pRK40 using the primers C31-8 (SEQ ID NO:8) and C31-9 (SEQ ID NO:9), which provide a

corrected Stop codon. The ends of this PCR fragment were digested with Eco47III and XhoI, the fragment was ligated into plasmid pRK40 and opened with Eco47III and XhoI to remove the fragment containing the defective stop codon. The resulting plasmid pRK55 contains the correct C31-Int gene as confirmed by DNA sequence analysis.

The C31-Int gene was isolated from pRK55 as 1.85 kb fragment by digestion with NotI and XhoI and ligated into the generic mammalian expression vector pRK50 (see below), opened with NotI and XhoI, giving rise to the C31-Int expression vector pCMV-C31-Int(wt). pCMV-C31-Int(wt) (SEQ ID NO:10) contains a 700 bp cytomegalovirus immediate early gene promoter (position 1 – 700), a 270 bp hybrid intron (position 701 – 970), the C31-Int gene (position 978 – 2819), and a 189 bp synthetic polyadenylation sequence (position 2831 – 3020).

For the construction of pCMV-C31-Int(NNLS) a 1.5 kb fragment was amplified by PCR from phage DNA using oligonucleotides C31-2 (SEQ ID NO:98) and C31-3 (SEQ ID NO:7). The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmid pRK41 (SEQ ID NO: 99). A 1100 bp fragment generated by digestion of plasmid pRK41 with NcoI and NotI was then ligated into plasmid pRK55 (SEQ ID NO:80), opened with NcoI and NotI, giving rise to the plasmid pRK63 (SEQ ID NO:81). The C31-Int gene with N-terminal NLS was isolated as a 1.8 kb fragment from pRK63 by digestion with NotI and XhoI and ligated into the mammalian expression vector pRK50, opened with NotI and XhoI, giving rise to the C31-Int expression vector pCMV-C31-Int(NNLS). pCMV-C31-Int(NNLS) (SEQ ID NO:73) contains a 700 bp Cytomegalovirus immediate early gene promoter (position 1 – 700), a 270 bp hybrid intron (position 701 – 970), the C31-Int gene with N-terminal NLS (position 976 – 2838), and a 189 bp synthetic polyadenylation sequence (position 2851 – 3040).

For the construction of pCMV-C31-Int(CNLS), the 3'-end of the C31-Int gene was amplified from pCMV-C31-Int(wt) as a 300 bp PCR fragment using the primers C31-8 (SEQ ID NO:8) and C31-2-2 (SEQ ID NO:11). Primer C31-2-2 modifies the 3'-end of the wildtype C31-Int gene such that the stop codon is replaced by a sequence of 21 basepairs coding for the SV40 T-antigen nuclear localisation sequence of 7 amino acids (Proline-Lysine-Lysine-Lysine-Arginine-Lysine-Valine) (Kalderon et. al, Cell, 39, 499 – 509 (1984)), followed by a new stop

codon. The ends of this 300 bp PCR fragment were digested with Eco47III and XhoI, the fragment was ligated into plasmid pCMV-C31-Int(wt) and opened with Eco47III and XhoI to replace the 3'-end of the wildtype C31-Int gene resulting in the plasmid pCMV-C31-Int(CNLS). The identity of the new gene segment was verified by DNA sequence analysis. pCMV-C31-Int(CNLS) (SEQ ID NO:12) contains a 700 bp cytomegalovirus immediated early gene promoter (position 12 – 711), a 270 bp hybrid intron (position 712 – 981), the modified C31-Int gene (position 989 – 2851), and a 189 bp synthetic polyadenylation sequence (position 2854 – 3043).

10

To generate the Cre expression plasmid pCMV-Cre (SEQ ID NO:13), the coding sequence of Cre recombinase (Sternberg et al., J. Mol. Biol., 187, 197 – 212 (1986)) with a N-terminal fusion of the 7 amino acid SV40 T-antigen NLS (see above) was recovered from plasmid pgk-Cre and cloned into the NotI and XhoI sites of plasmid pRK50. PRK50 (SEQ ID NO:14) is a generic expression vector for mammalian cells based on the cloning vector pNEB193 (New England Biolabs Inc, Beverly, MA, USA). PRK50 was built by insertion into pNEB193 of a 700 bp cytomegalovirus immediated early gene (CMV-IE) promoter (position 1-700) from plasmid pIREShyg (GenBank#U89672; Clontech Laboratories Inc, Palo Alto, CA, USA), a synthetic 270 bp hybrid intron (position 701-970), consisting of a adenovirus derived splice donor and an IgG derived splice acceptor sequence (Choi et al., Mol. Cell. Biol., 11, 3070 – 3074 (1991)), and a 189 bp synthetic polyadenylation sequence (position 1000-1188) build from the polyadenylation consensus sequence and 4 MAZ polymerase pause sites (Levitt et al., Genes&Dev., 3, 1019 – 1025 (1989); The EMBO J. 13, 5656 – 5667 (1994)). The positive control plasmid pRK64(ΔCre) (SEQ ID NO:15) was generated from pRK64 by transformation into the Cre expressing E. coli strain 294-Cre (Buchholz et al., Nucleic Acids Res., 24, 3118 – 3119 (1996)).

30 One of the transformed subclones was confirmed for the Cre mediated deletion of the loxP-flanked cassette by restriction mapping and further expanded. Plasmid pUC19 is a cloning vector without eukaryotic control elements used to equalise DNA amounts for transfections (GenBank#X02514; New England Biolabs Inc, Beverly, MA, USA). All plasmids were propagated in DH5 α E. coli cells (Life 35 Technologies GmbH, Karlsruhe, Germany) grown in Luria-Bertani medium and

purified with the plasmid DNA purification reagents "Plasmid-Maxi-Kit" (Qiagen GmbH, Hilden, Germany) or "Concert high purity plasmid purification system" (Life Technologies GmbH, Karlsruhe, Germany). Following purification, the plasmid DNA concentrations were determined by absorption at 260 nm and 280 nm in UVette cuvettes (Eppendorf-Netheler-Hinz GmbH, Hamburg, Germany) using a BioPhotometer (Eppendorf-Netheler-Hinz GmbH, Hamburg, Germany) and the plasmids were diluted to the same concentration; finally these were confirmed by separation of 10 ng of each plasmid on an ethidiumbromide-stained agarose gel.

10

B. Cell culture and transfections: Chinese hamster ovary (CHO) cells (Puck et al., J. Exp. Med., 108, 945 (1958)) were obtained from the Institute for Genetics (University of Cologne, Germany) as a population adapted to growth in DMEM medium. The cells were grown in DMEM/Glutamax medium (Life Technologies) 15 supplemented with 10% fetal calf serum at 37°C, 10% CO₂ in humid atmosphere and passaged upon trypsinisation. One day before transfection 10⁶ cells were plated into a 48-well plate (Falcon). For the transient transfection of cells with plasmids each well received into 250 ml of medium a total amount of 300 ng supercoiled plasmid DNA complexed before with the FuGene6 transfection 20 reagent (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturers protocol. Each 300 ng DNA preparation (Fig.2 sample 4 to 11) contained 50 ng of the luciferase expression vector pUHC13-1 (Gossen et al., Proc Natl Acad Sci USA., 89 5547-5551 (1992)), 50 ng of the substrate vector pRK64, 0.5 ng or 1 ng of one of the recombinase expression vectors pCMV- 25 C31Int(wt), pCMV-C31Int(NNLS), pCMV-C31Int(CNLS) or pCMV-Cre and 199 ng or 199.5 ng of pUC19 plasmid, except for the controls which received 50 ng of pUHC13-1 together with 50 ng of pRK64 (sample 3) or pRK64(Δcre) (sample 2) and 200 ng pUC19, or 50 ng pUHC13-1 with 250 ng pUC19 (sample 1). Transfections of Ssv and XisA recombinases (Fig.3) also contained 50 ng of the 30 luciferase expression vector pUHC13-1, 50 ng of substrate vectors pPGKattA and pPGKnif and 10 ng or 20 ng of recombinase expression vector pCMV-SSV or pCMV-SSV(NNLS) or 25 ng or 100 ng of expression vectors pCMV-XisA/ pCMV-XisA(NNLS). Plasmid pUC19 was added to a total amount of 300 ng plasmid DNA. As the C31-Int expression vectors are 15% larger in size than pCMV-Cre and the 35 same amounts of DNA of the three plasmids were used for transfection, the

samples with C31-Int vectors received 15% less plasmid molecules as compared to the samples with Cre expression vector. The β -galactosidase values from C31-Int transfected samples by this value were not corrected and thus is a slight underestimation of the calculated C31-Int activities. For each sample to be tested 5 four individual wells were transfected. One day after the addition of the DNA preparations each well received additional 250 ml of growth medium. The cells of each well were lysed 48 hours after transfection with 100 ml lysate reagent supplemented with protease inhibitors (Roche Diagnostics). The lysates were centrifuged and 20 ml were used to determine the β -galactosidase activities 10 using the β -galactosidase reporter gene assay (Roche Diagnostics) according to the manufacturers protocol in a Lumat LB 9507 luminometer (Berthold). To measure luciferase activity, 20ml lysate was diluted into 250ml assay buffer (50mM glycylglycin, 5mM MgCl₂, 5mM ATP) and the "Relative Light Units" (RLU) were counted in a Lumat LB 9507 luminometer after addition of 100 ml of a 1 15 mM luciferin (Roche Diagnostics) solution. The mean value and standard deviation of the samples was calculated from the β -galactosidase and luciferase RLU values obtained from the four transfected wells of each sample.

C. Results: To set up an assay system for the measurement of C31-Int and Cre 20 recombinase efficiency in mammalian cells the recombination substrate vector pRK64 shown in Figure 1E was first constructed. pRK64 contains a SV40 promoter region for expression in mammalian cells followed by a 1.1 kb cassette which consists of the coding region of the puromycin resistance gene and a polyadenylation signal sequence. This cassette is flanked at the 5'-end by the 84 25 bp attB and at the 3'-end by the 84 bp attP recognition site of C31-Int (Fig. 1 and 6). These attB and attP sites are located on the same DNA molecule and oriented in a way to each other which allows the deletion of the flanked DNA segment. The same orientation of attB and attP sites is used naturally by the ϕ C31 phage and the bacterial genome, leading to the integration of the phage 30 genome when both sites are located on different DNA molecules (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)). To measure C31-Int and Cre recombinase activities with the same substrate vector, pRK64 contains in addition two Cre recognition (loxP) sites in direct orientation next to the att sites. Since the att/lox-flanked cassette in plasmid pRK64 is inserted between the SV40 35 promoter and the coding region of the β -galactosidase gene, its presence inhibits

β-galactosidase expression as the SV40 promoter derived transcripts are terminated at the polyadenylation signal of the puromycin gene. Plasmid pRK64 is turned into a β-galactosidase expression vector upon C31-Int or Cre mediated deletion of the att/lox-flanked puromycin cassette since the remaining single att 5 and loxP site do not substantially interfere with gene expression.

For the expression of recombinases a mammalian expression vector was 10 designed which contains the CMV immediate early promoter followed by a hybrid intron, the coding region of the recombinase to be tested, and an artificial polyadenylation signal sequence. The backbone sequence of the four recombinase expression vectors shown in Figure 1A-D is identical to each other except for the recombinase coding region. Plasmid pCMV-C31Int(wt) (Fig. 1A) contains the nonmodified (wildtype) 1.85 kb coding region of C31-Int as found in the genome of phage ΦC31 (Kuhstoss, et al., J. Mol. Biol. 222, 897-908 (1991)). 15 Plasmid pCMV-C31Int(NNLS) (Fig. 1B) contains a modified C31-Int gene coding for the full length C31-Int protein with a N-terminal extension of 7 amino acids derived from the SV40 virus large T antigen which serves as a nuclear localisation signal (NLS). Plasmid pCMV-C31Int(CNLS) (Fig. 1C) contains a C-terminal extension of 7 amino acids derived from the SV40 virus large T antigen 20 which serves as a nuclear localisation signal (NLS). Plasmid pCMV-Cre (Fig. 1D) contains the 1.1 kb Cre coding region with an N-terminal fusion of the 7 amino acid NLS of the SV40 T-antigen. For Cre recombinase it has been shown that the N-terminal addition of the SV40 T-antigen NLS does not increase its 25 recombination efficiency in mammalian cells (Le et al., Nucleic Acids Res., 27, 4703 - 4709 (1999)).

As a test system to compare the efficiency of the 4 recombinases the same amount of plasmid DNA of each of the recombinase expression vectors together with a fixed amount of the reporter plasmid pRK64 was transiently introduced 30 into Chinese Hamster Ovary (CHO) cells. Thus, in this assay design the efficiency of the various recombinases on an extrachromosomal substrate introduced into the CHO cells was compared as a circular plasmid. Two days after transfection the cells from the various samples were lysed and the activity of β-galactosidase 35 in the lysates was determined by a specific chemiluminescence assay and expressed in "Relative Light Units" (RLU (β-Gal)) (Fig. 2). In addition all samples

contained a fixed amount of a luciferase expression vector to control for the experimental variation of cell transfection and lysis. For this purpose the lysates of each sample were also tested for luciferase activity with a specific chemiluminescence assay and the values expressed as "Relative Light Units"

5 (RLU (Luciferase)) (Fig. 2). All transfection samples contained in addition varying amounts of the unrelated cloning plasmid pUC19 so that all samples were equalised to the same amount of plasmid DNA. As a positive control for β -galactosidase a derivative of the recombination reporter pRK64 was used in which the loxP flanked 1.1 kb cassette has been removed through Cre mediated
10 recombination in *E. coli* giving rise to plasmid pRK64(Δ Cre). As negative controls served samples which received the unrecombined reporter plasmid pRK64 but no recombinase expression vector as well as samples set up with the pUC19 plasmid alone.

15 To determine the relative efficiency of the tested recombinases the RLU values of β -galactosidase were divided individually for each sample by the RLU values obtained for luciferase and multiplied with 10^5 . From the values of the four data points of each sample the mean value and standard deviation was calculated as an indicator of recombinase activity (Gal/Luc) (Fig. 2). The relative activity of the
20 tested recombinases was then compared to the positive control defined as an activity of 1.

As shown in Fig. 2, the expression of Cre recombinase (samples 10 and 11) resulted in a 150 to 170-fold increase of β -galactosidase activity as compared to
25 the negative control (sample 3), demonstrating the wide dynamic range of our test system. Each recombinase vector was tested using two different amounts of DNA for transfection (0.5 and 1ng/sample), which in the case of Cre resulted in 63% and 72% recombinase activity (samples 10 and 11 as compared to the positive control). These two values establish that the DNA amounts used are
30 close to the test systems saturation for recombinase expression as the doubling of DNA amounts resulted only in a minor increase of recombinase activity.

In comparison to Cre, the expression of wildtype C31-Int resulted in a considerably lower recombinase activity of 23% and 30% (Fig. 2, samples 4 and
35 5) as compared to the positive control. These values represent 37% and 42%

recombinase activity for wildtype C31-Int as compared to Cre recombinase (compare samples 4 and 5 with 10 and 11). Upon the expression of C31-Int fused with the N-terminal NLS (C31-Int(NNLS)) values of 32% and 36% recombinase activity (samples 6 and 7) were obtained as compared to the 5 positive control. The C31-Int(NNLS) values represent 51% and 50% recombinase activity as compared to Cre (compare samples 6 and 7 to 10 and 11). Thus, the activity of C31-Int in mammalian cells is just moderately enhanced by the addition of a NLS signal.

Surprisingly, upon the expression of C31-Int fused with the C-terminal NLS (C31-10 Int(CNLS)) values of 50% and 65% recombinase activity (samples 8 and 9) were obtained as compared to the positive control. The C31-Int(CNLS) values represent 79% and 90% recombinase activity as compared to Cre recombinase (compare samples 8 and 9 to 10 and 11). Unexpectedly, C31-Int(CNLS) exhibits a dramatic, more than twofold increase of recombinase activity in comparison to 15 C31-Int(wt) (compare samples 8 and 9 to 4 and 5).

In order to test whether the addition of a NLS sequence may be a general, simple method to enhance recombinase activity in mammalian cells we extended our studies by two additional recombinases: XisA recombinase (XisA) derived 20 from the cyanobacterium Anabaena, and SSV-Integrase (SSV-Int) derived from the SSV1 virus of the thermophilic bacterium Sulfolobus shibatae. To this end we constructed mammalian expression vectors for the wildtype forms of XisA and SSV recombinases and compared their activity to versions which were modified by the N-terminal addition of the 7 amino acid NLS of the SV40 T-antigen. These 25 recombinases were compared by the use of the reporter vector shown in Fig.1E, except that the att elements of C31-Int were replaced by the rif recognition sequences for XisA or the att sequences for SSV-Int. As described above for C31-Int, recombinase activities were tested by transient transfection into CHO cells using the reporter vector derived β -galactosidase activity as readout and 30 cotransfected luciferase as internal control.

As shown in Fig.3 for both, XisA and SSV recombinases the addition of a NLS sequence did not improve their activity in a mammalian cell line as compared to the wildtype forms. At both DNA concentrations tested wildtype XisA exhibits a significant recombination activity as compared to the reporter vector alone 35 (compare samples 2 and 3 to sample 1), but this activity is not altered by the

addition of an NLS (compare samples 2 and 3 to samples 4 and 5). SSV-Int exhibits only a low recombination activity (compare samples 7 and 8 with sample 6) which is also not enhanced by the addition of a NLS (compare samples 9 and 10 with samples 7 and 8). From these results we conclude that the addition of a 5 NLS to an inefficient recombinase is not a general, simple method to improve its performance in mammalian cells.

Taken together, in the transient transfection test system shown in Figure 2 a more than twofold activity increase of the Φ C31 Integrase could be achieved by 10 the C-terminal, but not the N-terminal addition of the SV40 T antigen NLS signal. As this signal sequence has been characterised to act as a nuclear localisation signal (Kalderon et. al, Cell, 39, 499 - 509 (1984)) we conclude that the efficiency increase of C31-Int(CNLS) is the result of the improved nuclear accumulation of this recombinase. The relative inefficiency of C31-Int (NNLS) 15 may be explained by the inaccessibility of the NLS peptide to the nuclear import machinery at the N-terminal position of the C31-Int protein.

In particular, it could be shown that C31-Int(CNLS) recombines extrachromosomal DNA in mammalian cells almost as efficient as the widely used Cre recombinase and thus provides an additional or alternative recombination 20 system of highest activity. The efficiency increase of C31-Int(CNLS) as compared to its wildtype form is regarded as an invention of substantial use for biotechnology.

Example 2

25 As demonstrated in example 1 C31-Int recombinase with the C-terminal fusion of the SV40 T-antigen NLS (C31-Int(CNLS)) shows in mammalian cells a recombination activity comparable to Cre recombinase on an extrachromosomal plasmid vector. It was further tried to test whether C31-Int(CNLS) exhibits a similar activity on a recombination substrate which is chromosomally integrated 30 into the genome of mammalian cells. This question is critical for the use of a recombination system for genome engineering as it is possible that a recombinase may act efficiently on extrachromosomal substrates but is impaired if the recognition sites are part of the mammalian chromatin. To characterise the recombination activity of C31-Int(CNLS) and C31-Int(NNLS) on a chromosomal 35 substrate the pRK64 reporter plasmid (Fig. 1E) was stably integrated, containing

a pair of loxP and att sites, into the genome of a mammalian cell line. One of the stable transfected clones was chosen for further analysis and was transiently transfected with recombinase expression vectors coding for C31-Int(CNLS), C31-Int(NNLS), C31-Int(wt) or Cre recombinase. The activity of β -galactosidase derived from the Cre expression vector recombined in these cells was taken as a measure of recombination efficiency.

A. Plasmid constructions: all plasmids used and their purification are described in example 1.

10

B. Cell culture and transfections: To generate a stably transfected C31-Int reporter cell line 2.5×10^6 NIH-3T3 cells (Andersson et al., Cell, 16, 63-75 (1979); DSMZ#ACC59; DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany)

15

were electroporated with 5 μ g pRK64 plasmid DNA linearised with ScaI and plated into 10cm petri dishes. The cells were grown in DMEM/Glutamax medium (Life Technologies) supplemented with 10% fetal calf serum at 37°C, 10% CO₂ in humid atmosphere, and passaged upon trypsinisation. Two days after transfection the medium was supplemented with 1mg/ml of puromycin (Calbiochem) for the 20 selection of stable integrants. Upon the growth of resistant colonies these were isolated under a stereomicroscope and individually expanded in the absence of puromycin. To demonstrate stable integration of the transfected vector, genomic DNA of puromycin resistant clones was prepared according to standard methods and 5-10 μ g were digested with EcoRV. Digested DNA was separated in a 0.8% 25 agarose gel and transferred to nylon membranes (GeneScreen Plus, NEN DuPont) under alkaline conditions for 16 hours. The filter was dried and hybridised for 16 hours at 65°C with a probe representing the 5' part of the E. coli β -galactosidase gene (1.25 kb NotI - EcoRV fragment of plasmid CMV- β -pA (R. Kühn, unpublished). The probe was radiolabelled with P32-marked α -dCTP 30 (Amersham) using the Megaprime Kit (Amersham). Hybridisation was performed in a buffer consisting of 10% dextran sulfate, 1% SDS, 50 mM Tris and 100 mM NaCl, pH7.5). After hybridisation the filter was washed with 2x SSC/1%SDS and exposed to BioMax MS1 X-ray films (Kodak) at ~ 80°C.

Transfections of the selected clone 3T3(pRK64)-3 with plasmid DNAs and the 35 measurement of β -galactosidase activities in lysates were essentially performed

as described in example 1 for CHO cells, except that 32ng or 64ng of the recombinase expression plasmids and 218 or 186 ng of pUC19 plasmid were used and the pRK64 plasmid was omitted from all samples.

5 C. Histochemical detection of β -galactosidase activity in transfected 3T3(pRK64)-3 cells

To directly demonstrate β -galactosidase expression in recombinase transfected cells, 10^4 3T3(pRK64)-3 cells were plated one day before transfection into each well of a 48-well tissue culture plate (Falcon). For the transient transfection of 10 cells with plasmids each well received into 250 μ l of medium a total amount of 150 ng supercoiled plasmid DNA complexed before with the FuGene6 transfection reagent (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturers protocol. Each 150 ng DNA preparation contained 50 ng of the recombinase expression vector pCMV-Cre or pCMV-C31Int(CNLS) and 100ng of 15 the pUC19 plasmid. After 2 days the culture medium was removed from the wells, the wells were washed once with phosphate buffered saline (PBS), and the cells were fixed for 5 minutes at room temperature in a solution of 2% formaldehyde and 1% glutaraldehyde in PBS. Next, the cells were washed twice with PBS and finally incubated in X-Gal staining solution for 24 hours at 37°C 20 (staining solution: 5 mM $K_3(Fe(CN)_6$), 5 mM $K_4(Fe(CN)_6$), 2 mM $MgCl_2$, 1mg/ml X-Gal (BioMol) in PBS) until photographs were taken.

D. Results

To generate a mammalian cell clone with a stable genomic integration of the 25 C31-Int and Cre recombinase reporter plasmid pRK64, the murine fibroblast cell line NIH-3T3 was electroporated with linearised pRK64 DNA (Fig.1D; see also example 1) and subjected to selection in puromycin containing growth medium. Plasmid pRK64 contains in between the pair of loxP and att sites the coding 30 region of the puromycin resistance gene expressed from the SV40-IE promoter. Thirty-six puromycin resistant clones were isolated and the genomic DNA of 19 clones was analysed for the presence and copy number of the pRK64 DNA. Three clones, which apparently contain 2 - 4 copies of pRK64, were further characterised on the single cell level for the expression of β -galactosidase upon transient transfection with the Cre expression vector pCMV-Cre (Fig. 1C). The cell 35 clone with the largest proportion of β -galactosidase positive cells, 3T3(pRK64)-3,

was selected as most useful for the planned studies on C31-Int and Cre recombinase efficiency.

To compare the efficiency of wildtype C31-Int (C31-Int(wt)), C31-Int(CNLS),
5 C31-Int(NNLS), and Cre recombinases 32ng or 64 ng of the recombinase expression vectors pCMV-C31Int(wt), pCMV-C31Int(CNLS), pCMV-C31Int(NNLS), or pCMV-Cre (Fig. 1 A-D) together with the unrelated cloning plasmid pUC19 were transiently introduced into 3T3(pRK64)-3 cells, such that all samples contained the same amount of plasmid DNA. As a negative control a sample
10 prepared with the pUC19 plasmid alone was used. Two days after transfection the cells from the various samples were lysed and the activity of β -galactosidase in the lysates was determined by a specific chemiluminescence assay and expressed in "Relative Light Units" (RLU)(β -Gal) (Fig. 4). From the values of the four data points of each sample the mean value and standard deviation was
15 calculated as an indicator of recombinase activity (Fig.4). The relative activity of the tested recombinases was then compared to the highest value obtained with the Cre expression vector, defined as an activity of 1.

As shown in Figure 4 the expression of Cre recombinase (samples 8 and 9)
20 resulted in a 36 to 49-fold increase of β -galactosidase activity as compared to the negative control (sample 1), demonstrating the dynamic range of the test system used. Each recombinase vector was tested using two different amounts of DNA for transfection (32 ng and 64 ng/sample), which in the case of Cre resulted in 73% and 100% recombinase activity (samples 8 and 9). These two values
25 establish that the DNA amounts used are not far from the linear scale of the test systems ability for recombinase expression as the twofold increase of the amount of DNA also resulted in a significant increase of recombinase activity.

The expression of wildtype C31-Int (Fig. 4, samples 2 and 3) resulted in a low
30 recombinase activity of 4% and 10% as compared to the values obtained by Cre transfection (compare samples 2 and 3 with 8 and 9). This activity was only moderately enhanced by the expression of C31-Int(NNLS) to values of 19% and 22% of Cre activity (compare samples 4 and 5 with samples 8 and 9). Upon the expression of C31-Int(CNLS) values of 48% and 78% recombinase activity were
35 obtained as compared to Cre recombinase (compare samples 6 and 7 to 8 and

9). Hence, C31-Int(CNLS) exhibits an 12-fold higher activity than C31-Int(wt) at 32 ng plasmid DNA (Fig.4, compare samples 6 and 2) and an 8-fold higher activity than C31-Int(wt) at 64 ng plasmid DNA (Fig.4, compare samples 7 and 3).

5

In addition, it was aimed to directly demonstrate *in situ* the expression of β -galactosidase in 3T3(pRK64)-3 cells after transfection with Cre or C31-Int(CNLS) recombinase plasmid. Two days after transfection the cells were fixed *in situ* and incubated with the histochemical X-Gal assay which detects β -galactosidase positive cells by a blue precipitate. As shown in Figure 5 stained cells were found at a comparable frequency in the samples transfected with the Cre or C31-Int(CNLS) expression vectors but not in the nontransfected control. This result confirms that the β -galactosidase activities measured by chemiluminescence upon recombinase transfection (Fig. 4) results from a population of individual, 10 recombinant reporter cells.

15 In conclusion, upon the transient transfection of recombinase expression vectors into a cell line with a genomic integration of the recombination substrate vector, a 8 ~ 12-fold activity increase of the Φ C31 Integrase by the C-terminal fusion with the SV40 T-antigen NLS signal was found. As this signal sequence has been 20 characterised to act as a nuclear localisation signal (Kalderon et. al, Cell, 39, 499 ~ 509 (1984)), it was concluded that the dramatic efficiency increase of C31-Int(CNLS) is the result of the improved nuclear accumulation of this recombinase. The approximately tenfold activity increase of C31-Int(CNLS) upon 25 expression in a cell line with a genomic integration of the substrate vector is considerably higher than the activity increase found upon the transient expression of both vectors (see example 1). Thus, a substrate vector integrated into the chromatin of a mammalian cell may pose more stringent requirements 30 on recombinase activity to be recombined as compared to an extrachromosomal substrate.

The dramatic activity increase of C31-Int(CNLS), as compared to its wildtype form, on a stable integrated substrate in mammalian cells is an invention of significant practical use as this recombinase is as efficient as the widely used

Cre/loxP system; thus, C31-Int(CNLS) provides an additional or alternative recombination system of highest activity.

Example 3

5 To demonstrate that the increase in β -galactosidase activity obtained by the cotransfection of a C31-Int expression vector and the reporter vector pRK64 into mammalian cells is in fact the result of recombinase mediated deletion, one of the recombination products was detected by a specific polymerase chain reaction (PCR). The amplified PCR product was cloned and its sequence determined. The 10 obtained sequence confirms that C31-Int mediated deletion of the test vector occurs in a mammalian cell line and that the recombination occurs at the known breakpoint within the attB and attP sites.

15 **A. Plasmid constructions:** The construction of plasmids pRK64, pCMV-Cre and pCMV-C31-Int(wt) is described in Example 1. To simulate the recombination of pRK64 by C31-Int, the sequence between the CAA motives of the att sites (boxed in Fig.5) was deleted from the computerfile of pRK64, giving rise to the sequence of pRK64(Δ Int) (SEQ ID NO:16).

20 **B. Transfection of Cells and PCR amplification:** MEF5-5 mouse fibroblasts (Schwenk et al., 1998) (20000 cells per well of a 12 well plate (Falcon)) were transfected with 0.5 μ g pRK64 alone or together with 250 ng pCMV-Int(wt) or pCMV-Cre using the FuGene6 transfection reagent following the manufacturers protocol (Roche Diagnostics). After 2 days DNA was extracted from these cells 25 according to standard methods and used for PCR amplification with Primers P64-1 (SEQ ID NO:17; complementary to position 111-135 of pRK64(Δ Int)) and P64-4 (SEQ ID NO:18; complementary to position 740-714 of pRK64(Δ Int)) using the Expand High Fidelity PCR kit (Roche Diagnostics). PCR products were separated on a 0.8% agarose gel, extracted with the QuiaEx kit (Quiagen) and cloned into 30 the pCR2.1 vector using the TA cloning kit (Invitrogen) resulting in plasmid pRK80d. The sequence of its insert, seq80d (SEQ ID NO:19), was determined using the reverse sequencing primer and standard sequencing methods (MWG Biotech).

35 For the measurement of β -galactosidase activity the cells were lysed 2 days after

transfection and the β -galactosidase activities were determined with the β -galactosidase reporter gene assay (Roche Diagnostics) as described in example 1.

5 **C. Results:** As a test vector for C31-Int mediated DNA recombination plasmid pRK64 was used, which contains the 1.1 kb coding region of the puromycin resistance gene flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site of C31-Int (Fig. 5). These attB and attP sites are located on the same DNA molecule and oriented in a way to each other which allows the
10 deletion of the att-flanked DNA segment. The same orientation of attB and attP sites is used naturally by the ϕ C31 phage and the bacterial genome for the integration of the phage genome when both sites are located on different DNA molecules (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)). As a positive control, vector pRK64 contains in addition two Cre recombinase
15 recognition (loxP) sites in direct orientation next to the att sites. Since the att-flanked DNA segment in plasmid pRK64 is inserted between a promoter active in mammalian cells and the β -galactosidase gene, its deletion can be measured by the increase of β -galactosidase activity. The expected product of C31-Int mediated deletion of plasmid pRK64 is shown in Fig. 6, designated as
20 pRK64(Δ Int). If the recombination between attB and attP occurs as described in bacteria (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 – 5510 (1998)), a single attR site is generated and left on the parental plasmid (Fig. 6) while the flanked DNA is excised and contains an attL site. Beside the measurement of β -galactosidase activity, C31-Int mediated recombination of pRK64 can be directly
25 detected on the DNA level by a specific polymerase chain reaction (PCR) using the primers P64-1 and P64-4 (Fig. 6). These primers, located 5' of the attB site (P64-1) and 3' of the attP site, are designed to amplify a PCR product of 630 bp length upon the C31-Int mediated recombination of pRK64. For the expression of C31-Int in mammalian cells plasmid pCMV-C31(wt) was used, which contains the
30 CMV-IE-Promoter upstream of the C31-Int coding region followed by a synthetic polyadenylation signal (see Example 1 and Fig.1).

The recombination substrate vector pRK64 was transiently transfected into the murine fibroblast cell line MEF5-5 either alone, or together with the C31-Int expression vector pCMV-C31(wt), or together with an expression vector for Cre recombinase, pCMV-Cre. Two days after transfection half the cells of each sample

was lysed and used to measure β -galactosidase activity by chemiluminescence, and the other half was used for the preparation of DNA from the transfected cells for PCR analysis. The β -galactosidase levels of the 3 samples were found as following (expressed as "Relative Light Units" (RLU) with standard deviation (SD) of the β -galactosidase assay):

<u>Sample</u>	<u>RLU (SD)</u>
1) pRK64	692 \pm 5
2) pRK64 + pCMV-Cre	8527 \pm 269
10) 3) pRK64 + pCMV-C31(wt)	1288 \pm 93

As the coexpression of the test vector pRK64 together with the C31-Int expression vector in sample 3 leads to a significant increase of β -galactosidase activity as compared to pRK64 alone, this result suggests that pRK64 is 15 recombined by C31-Int as anticipated in Fig. 6.

Next, cellular DNA was prepared from the three samples and tested for the occurrence of the expected Cre or C31-Int generated deletion product by PCR using primers P64-1 and P64-4 for amplification. As shown in Fig. 7 an 20 amplification product of the expected size was found only in the samples cotransfected with the Cre or C31-Int recombinase expression vectors (Fig. 7A, lane 3 and lane 4). The PCR products amplified from pRK64 recombined by C31-Int or Cre are of the same size but should be recombined via the attB/P or loxP sites, respectively.

25 To prove that the PCR product found after cotransfection of plasmids pRK64 and pCMV-C31(wt) represents in fact the deletion product of C31-Int mediated recombination, this DNA fragment was cloned into a plasmid vector and its DNA sequence determined. One clone, pRK80d, was analysed, and its sequence showed exactly the sequence of an attR site as expected from C31-Int mediated 30 deletion of pRK64 (Fig. 7B, compare to Fig. 6).

In conclusion, this experiment demonstrates that C31-Int mediated deletion of a 35 vector containing a pair of attB/attP sites occurs in a mammalian cell line, and that the recombination occurs within the same 3 bp breakpoint region of attB and attP as found in bacteria (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 -

5510 (1998)). Thus, it was concluded that an increase of β -galactosidase activity observed by cotransfection of the pRK64 reporter vector and a C31-Int expression vector in mammalian cells truly reflects C31-Int recombinase activity.

5

Example 4

As has been demonstrated in examples 1-3, the C31-Int recombinase with the C-terminal fusion of the SV40 T-antigen NLS (C31-Int(CNLS)) shows a 10 recombination activity comparable to Cre recombinase on an extrachromosomal as well as a chromosomally integrated target in mammalian cells in vitro. To test whether C31-Int(CNLS) exhibits activity in mice, transgenic mice carrying a C31-15 Int(CNLS) expression vector were generated. These transgenic mice were crossed with reporter mice carrying the recombinase substrate. Recombination-mediated expression of β -galactosidase, which can be measured by staining with the substrate X-Gal, was analyzed in testes of double transgenic progeny carrying both the recombinase and the reporter.

20 A. Plasmid constructions: For the construction of the C31-Int(CNLS) transgene expression vector, the C31Int gene with C-terminal NLS was isolated as a 2 kb-fragment generated by restriction of pCMV-C31Int(CNLS) (SEQ ID NO: 12) with BglII. The fragment was ligated into the BglII restriction site of the vector pCAGGS-Cre-pA (SEQ ID NO:104) giving rise to the plasmid pCAGGS-C31CNLS-pA (SEQ ID NO:105). In pCAGGS-C31CNLS-pA the C31-Int(CNLS) (position 25 1891-3753) is transcribed from the CAGGS promoter (position 1-1616) and followed by the SV40 late region polyadenylation sequence (position 3763-3941).

30 B. Production of transgenic mice: For the embryo injections a 3.95 kb-fragment was generated by restriction of the plasmid pCAGGS-C31CNLS-pA with PstI and AscI. This fragment was purified as follows: DNA bands were separated on an agarose-gel without ethidiumbromide. One part of the gel was stained with ethidiumbromide to locate the band to excise. The DNA was electroeluted from the excised band with S&S Biotrap Elution Chamber in 1x TAE (40 mM Tris-acetate, 1 mM EDTA) overnight. The DNA was precipitated from the eluate with 35 1/10 volume 3M sodium acetate and 2.5 volumes ethanol at -20 °C for several

hours. The DNA was pelleted by centrifugation at 13000 rpm for 30 min and washed twice with 70 % ethanol. The dried DNA pellet was resuspended in TE (10 mM Tris, 1 mM EDTA, pH 8). Subsequently the precipitation procedure was repeated once and the DNA resuspended in injection buffer (10 mM Tris pH 7.2, 5 0.1 mM EDTA). The sample was dialysed with Slide-A-Lyse Mini Dialysis Unit (Pierce) in injection buffer with several changes of buffer at 4°C overnight. Different amounts of the sample were checked on a gel to determine concentration. To generate transgenic mice, 5-10 fg of the purified fragment were injected into one pronucleus of (B6CBA)F2 mouse one-cell embryos. The 10 injected embryos were subsequently transferred into the oviduct of 0.5 day pseudopregnant NMRI females.

C. Analysis of transgenic mice: Mice were analyzed for the presence of the pCAGGS-C31CNLS-pA transgene by PCR using tail DNA and the primers C31- 15 screen 1 (SEQ ID NO:100) and C31-screen 2 (SEQ ID NO:101) amplifying a fragment of 500 bp. The PCR reaction contained 5 µl PCR buffer (Invitrogen), 2 µl 50 mM MgCl₂, 1.5 µl 10 mM dNTP-mix, 2 µl (10 pmol) of each primer, 0.5 µl Taq-polymerase (5 U/ µl) and water to a volume of 50 µl. The program used for the PCR reactions was: 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min in 30 20 cycles.

D. Analysis of C31-Int(CNLS) activity: Founder mice transgenic for the pCAGGS-C31CNLS-pA transgene were crossed to heterozygous C31 reporter mice carrying the C31 reporter construct in the ROSA26 locus (SEQ ID NO:106) (Fig. 8). 25 Offspring of the crosses were genotyped for the presence of the pCAGGS-C31CNLS-pA transgene by the PCR assay described in section C as well as for the ROSA26-C31 reporter allele by a LacZ-specific PCR assay. The PCR was performed using tail DNA and the primers β-Gal 3 (SEQ ID NO:102) and β-Gal 4 (SEQ ID NO:103) amplifying a fragment of 315 bp. The PCR reaction contained 5 µl PCR buffer (Invitrogen), 2.5 µl 50 mM MgCl₂, 2 µl 10 mM dNTP-mix, 1 µl (10 pmol) of each primer, 0.4 µl Taq-polymerase (5 U/ µl) and water to a volume of 30 50 µl. The program used for the PCR reactions was: 94 °C for 1 min, 60 °C for 1 min and 72 °C for 1 min in 30 cycles.

Testes from mice carrying the pCAGGS-C31CNLS-pA transgene as well as the 35 reporter locus and from a control mouse carrying the reporter allele only were

dissected. The tissues were imbedded in OCT Tissue freezing medium (Leica/Jung) and frozen in liquid nitrogen. Cryosections were generated from the embedded tissues using a Leica CM3050 cryomicrotome, dried on polylysine-coated slides for 1-4 hours and then stained as follows: Sections were fixed in 5 0.2 % glutaraldehyde, 5 mM EGTA, 2 mM MgCl₂ in 0.1 M PB (K₂HPO₄/ KH₂PO₄, pH 7.3) for 5 min at room temperature and washed in wash buffer (2 mM MgCl₂, 0.02 % Nonidet-40 in PB in 0.1 M PB) 3 times for 15 min. Then sections were stained in X-Gal-solution (0.6 mg/ ml X-Gal in DMSO, 5 mM potassium hexacyanoferrat III, 5 mM potassium hexacyanoferrat II in LacZ wash buffer) 10 overnight at 37 °C. After staining sections were washed in 1x PBS twice for 5 min. Dehydration was performed by washing the sections first with 70 %, 96 % and 100 % ethanol for 2 min each, then with a 1:1 mix of ethanol and xylol for 5 min and in the end only with xylol for 5 min. Before taking pictures sections were mounted in Entellan.

15

E. Results: To identify transgenic founder mice carrying the pCAGGS-C31CNLS-pA transgene, 29 mice born from the injection experiment were analyzed for the presence of the transgene. 5 founder mice (3 females and 2 males) were 20 identified. To analyze the activity of the C31-Int(CNLS) recombinase in transgenic mice, 2 of the female founder mice were crossed to heterozygous C31 reporter mice carrying a C31 reporter construct in the ROSA26 locus (Fig. 8). From each of these crosses, one offspring carrying the pCAGGS-C31CNLS-pA transgene as well as the C31 reporter allele was sacrificed. In order to determine 25 whether pCAGGS-C31CNLS-pA transgenic mice are able to delete an attB/P flanked DNA sequence in the mouse germline, tissue sections from the testes of the sacrificed animals were prepared and stained for β-galactosidase activity with X-Gal. Fig. 9 shows the result of the staining experiment for one of these mice (A) as well as a control mouse carrying only the reporter allele, but lacking the 30 pCAGGS-C31CNLS-pA transgene (B). Clear staining can be detected in the maturing sperm cells in about 50% of the tubules with the proportion of β-galactosidase expressing cells ranging between 10 and 100. No staining could be detected for the control mouse. This clearly demonstrates that C31-int-mediated recombination has taken place during spermatogenesis in the pCAGGS-C31CNLS-pA transgenic mice. These results show that the C31-int is functional *in vivo*, in a

transgenic mouse system and therefore provides a new tool to introduce specific deletions, inversions or integrations into the mouse germline.

Claims

1. A fusion protein comprising

(a) a recombinase domain comprising a recombinase protein or a mutant thereof having a recombinase activity similar to that of the corresponding wild-type recombinase and
(b) a signal peptide domain linked to said recombinase domain which directs nuclear import of said fusion protein in eucaryotic cells.

10 2. The fusion protein of claim 1, wherein the activity of the fusion protein in eucaryotic cells is significantly higher as compared to that of the wild-type recombinase corresponding to the recombinase of the recombinase domain.

15 3. The fusion protein of claim 1 or 2, wherein the recombinase domain comprises a recombinase protein belonging to the family of large serine recombinases or a mutant thereof, preferably the recombinase domain comprises a recombinase protein selected from the group consisting of bacteriophage Φ C31 integrase (C31-Int), coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase, CisA recombinase, XisF recombinase, transposon Tn4451 20 TnpX recombinase and lactococcal bacteriophage TP901-1 recombinase, or a mutant thereof; most preferably the recombinase protein is a C31-Int protein or a mutant thereof.

25 4. The fusion protein of claim 3, wherein the recombinase protein comprises a C31-Int having the amino acid sequence shown in SEQ ID NO:21 or a C-terminal truncated form thereof, said truncated form of the C31-Int preferably comprising amino acid residues of 306 to 613 of SEQ ID NO:21.

30 5. The fusion protein according to any one of claims 1 to 4, wherein the signal peptide domain is derived from yeast GAL4, SKI3, L29 or histone H2B proteins, polyoma virus large T protein, VP1 or VP2 capsid protein, SV40 VP1 or VP2 capsid protein, adenovirus E1a or DBP protein, influenza virus NS1 protein, hepatitis virus core antigen or the mammalian lamin, c-myc, max, c-myb, p53, c-erbA, jun, Tax, steroid receptor or Mx proteins, SV40 T-antigen or other proteins with

known nuclear localisation, preferably the signal peptide domain comprises a peptide which is derived from the SV40 T-antigen.

6. The fusion protein according to any one of claims 1 to 5, wherein the signal
5 peptide domain

(i) has a length of 5 to 74, preferably 7 to 15 amino acid residues, and/or
(ii) comprises a segment of 6 amino acid residues having at least 2 positively charged basic amino acid residues, said basic amino acid residues being preferably selected from lysine, arginine and histidine.

10

7. The fusion protein of claim 5 or 6, wherein the signal peptide domain comprises a peptide selected from a sequence shown in SEQ ID NOs:24 to 53, preferably the signal peptide comprises the amino acid sequence Pro-Lys-Lys-Lys-Arg-Lys-Val (SEQ ID NO:53).

15

8. The fusion protein according to any one of claims 1 to 6, wherein

(i) the signal peptide domain is linked to the N-terminal or C-terminal of the recombinase domain or is integrated into the recombinase domain, preferably the signal peptide domain is linked to the C-terminal of the recombinase domain;
20 and/or

(ii) the signal peptide domain is linked to the recombinase domain directly or through a linker peptide, said linker preferably having 1 to 30 essentially neutral amino acid residues.

25 9. The fusion protein of claim 1 comprising the amino acid sequence shown in SEQ ID NO:23.

10. A DNA coding for the fusion protein according to any one of claims 1 to 9.

30 11. A vector containing the DNA as defined in claim 10.

12. A microorganism containing the DNA of claim 10 and/or the vector of claim 11.

13. A process for preparing the fusion protein as defined in any one of claims 1 to 9 which comprises culturing a microorganism as defined in claim 11 under conditions suitable for expression of said fusion protein and recovering said fusion protein.

5

14. Use of the fusion protein as defined in any one of claims 1 to 9 to recombine DNA molecules, which contain recombinase recognition sequences for the recombinase protein of the recombinase domain, in eucaryotic cells.

10 15. A cell, preferably a mammalian cell containing the DNA sequence of claim 10 in its genome.

16. The cell of claim 15, also containing recognition sequences for the recombinase protein of the recombinase domain in its genome.

15

17. Use of the cell of claim 15 or 16 for studying the function of genes and for the creation of transgenic organisms.

20 18. A transgenic organism, preferably a transgenic non-human mammal containing the DNA sequence of claim 10 in its genome.

19. The transgenic organism of claim 18 also containing recognition sequences for the recombinase protein of the recombinase domain in its genome.

25 20. Use of the transgenic organism of claim 18 or 19 for studying gene function at various developmental stages.

30 21. A method for recombining DNA molecules of cells or organisms containing recombinase recognition sequences for the recombinase protein of the recombinase domain as defined in claims 1 to 9, which method comprises supplying the cells or organisms with a fusion protein as defined in claims 1 to 9 or with a DNA sequence of claim 10 and/or a vector of claim 11 which are capable of expressing said fusion protein in the cell or organism.

35 22. A method for recombining a DNA molecule containing recognition sequences

for a recombinase protein in a eucaryotic cell, said method comprising contacting the cell with a fusion protein according to claim 1 that recognizes said recognition sequences, wherein the fusion protein catalyzes recombination of the DNA molecule.

5

23. The fusion protein according to any one of claims 1 to 9 which catalyzes recombination at recognition sequences for the recombinase protein.

24 A transgenic organism, preferably a transgenic non-human mammal,

10 comprising a cell containing a DNA sequence coding for a recombinase fusion protein as defined in claims 1 to 9 and 23 in its genome.

SEQUENCE LISTING

5 <110> Artemis Pharmaceuticals GmbH
<120> Modified Recombinase
<130> 012787wo/JH/ml
10 <140>
<141>

<160> 108
15 <170> PatentIn Ver. 2.1

<210> 1
<211> 86
<212> DNA
20 <213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer C31-4
25 <400> 1
cgtgacggtc tcgaagccgc ggtgcgggtg ccagggcgtg cccttgggt ccccgccgc 60
gtactccacc tcaccatct ggtcca 86

30 <210> 2
<211> 86
<212> DNA
<213> Artificial Sequence

35 <220>
<223> Description of Artificial Sequence: primer C31-5

<400> 2
cgtggaccag atgggtgagg tggagtacgc gccccggag cccaaaggca cgccctggcc 60
cacgcaccgc ggcttcgaga ccgtca 86

45 <210> 3
<211> 90
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer C31-6
50 <400> 3
gatccagaag cggttttcgg gagtagtgcc ccaactgggg taaccttga gttctctcag 60
ttggggcggt agggtcgccg acatgacacg 90

55 <210> 4
<211> 90
<212> DNA
<213> Artificial Sequence
60 <220>
<223> Description of Artificial Sequence: primer C31-7-2

<400> 4
gatccgtgtc atgtcggcga ccctacgccc ccaactgaga gaactcaaag gttacccag 60
ttggggcact actcccgaaa accgcttctg 90

<210> 5
 <211> 7438
 <212> DNA
 5 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: vector pRK64

10 <400> 5
 cgtcatcacc gaaacgcgcg aggcaagtgt ggaatgtgt tcagtttaggg tggaaaagt 60
 ccccagggtc cccagcaggc agaagtagc aaagcatgc tctcaattag tcagcaacca 120
 ggtcccccaag caggcagaag tggcaaaagc atgcatacata attagtcagg aaccatagtc 180
 15 cccggccctaa ctccgccccat cccggccctaa actccgcggca gttccgcggca ttctccgccc 240
 catggctgac taatttttt tattttatgca gaggccgagg cccgcctcgcc ctaggaacag 300
 tcgacgacac tgcagagacc tacttcacta acaaccggta cagttcgtgg accagatggg 360
 tgagggtggag tacgcgcggc gggagcccaa gggcacgccc tggcacccgc accgcggctt 420
 20 cgagaccgtc acgaataact tcgtatagca tacattatac gaagtataaa gttgcatacg 480
 ctgcagggtcg gccgccacga cccggccggcc ggtgcgcggca ccattccctg acccacgcgg 540
 ctgacccctc acaaggagac gacccgcgac gacgtccccc gggccgtacg caccctcgcc gccgcgttcg ccgactaccc 600
 cggccacgcg caccacgtcg actcttcctc acgcgcgtcg ggctcgacat cggcaagggtg tgggtcgcgg aegacggcgc 660
 25 cggcggtggcg gtctggacca cggcccgccg atggcccgagt cccctcgcc gccgcgttcg ccgactaccc 720
 cctctggcg cccgacccac caggcgaagg cggcgccggc ggggtggccg 1020
 30 cgagcgcgcg ggggtggccg cggccggcgc ggcttcaccc tcaccggcga cgtcgagtgc cccggccaccc tcccttctta 1080
 gtgcatacgcc ccaaggaccc cggccggcgc ggcttcaccc tcaccggcga cgtcgagtgc cccggccaccc tcccttctta 1140
 aaggagcgcg cggccggcgc ggcttcaccc tcaccggcga cgtcgagtgc cccggccaccc tcccttctta 1200
 caccggcccg cggccggcgc ggcttcaccc tcaccggcga cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1260
 35 aggtttact tgctttaaaa aacccggcccg cggccggcgc ggcttcaccc tcaccggcga cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1320
 atgcataattt tgctttttttt aacccggcccg cggccggcgc ggcttcaccc tcaccggcga cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1380
 40 gcatcacaaa tttcacaaat aacccggcccg cggccggcgc ggcttcaccc tcaccggcga cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1440
 aactcatcaa ttttgcataat aacccggcccg cggccggcgc ggcttcaccc tcaccggcga cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1500
 aactgagaga actcaaagggt tccataactt cgtatagcat tttacaacgt cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1560
 45 tccataactt cgtatagcat tttacaacgt cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1620
 agcttggcac tggccgtcgt tttacaacgt cgtacttgc aacccggaccc gggggggcccc gccgaccccg 1680
 cttaatcgcc ttgcagcaca tccccccttc gccagctggc gtaatagcga agaggccgc 1740
 accgatcgcc citcccaaca gttgcgcagc ctgaatggcg aatggcgctt tgcctggtt 1800
 cccgcaccag aagcggtgcc gggaaaggctgg ctggagtgcg atcttcctga ggccgatact 1860
 50 gtcgtgtcc cctccaaactg gccagatgcac ggttacatgc cggcccatctt caccacgtca 1920
 acctatccca ttacggtcaa tccggcggtt gttcccacgg agaatccgc ggggttggta 1980
 55 tccgtcgat ttaatgttgc tggatggcg tggatggcg aatattttttt 2040
 tccgtcgat ttaatgttgc tggatggcg tggatggcg aatattttttt 2100
 gatggcgat tggatggcg tggatggcg tggatggcg aatattttttt 2160
 gacagtgcgtt tggatggcg tggatggcg tggatggcg aatattttttt 2220
 60 tccgtcgat tggatggcg tggatggcg tggatggcg aatattttttt 2280
 ctcgcgtgtt tggatggcg tggatggcg tggatggcg aatattttttt 2340
 65 gatggatggcg tggatggcg tggatggcg tggatggcg aatattttttt 2400
 gttcagatgt gggcgagtt gggcgagtt cttttttttt atggcagggt 2460
 gaaacgcagg tggatggcg tggatggcg tggatggcg aatattttttt 2520
 gtttatgcgg atcgctcactt gttttttttt atggcagggt 2580
 gaaatcccga atctctatcg gttttttttt atggcagggt 2640
 70 gaagcagaag cttttttttt atggcagggt 2700
 ctgaacggca agccgttgc tttttttttt atggcagggt 2760
 gtcaggtca tggatggcg tggatggcg tggatggcg aatattttttt 2820
 tttaaacgcgg tggatggcg tggatggcg tggatggcg aatattttttt 2880
 cgctacggcc tggatggcg tggatggcg tggatggcg aatattttttt 2940
 75 aatcgatgtt tggatggcg tggatggcg tggatggcg aatattttttt 3000
 gtgcagcgcg atcgatgtt tggatggcg tggatggcg aatattttttt 3060
 cacggcgctt atcaccgacgc tggatggcg tggatggcg aatattttttt 3120
 gtgcagttatg aaggccggcg tggatggcg tggatggcg aatattttttt 3180
 80 gcccggccgtt gttttttttt atggcagggt 3240
 ctttcgttac tggatggcg tggatggcg tggatggcg aatattttttt 3300
 aacagtcttgc ggggtttcgc tggatggcg tggatggcg aatattttttt 3360
 gggcgatcg tctggactt gggatggatcg tggatggatcg aatattgtca 3420

ccgtggtcgg cttacggcgg tgattttggc gatacggcga acgatcgcca gttctgtatg 3480
 aacggtctgg tctttggcga cccgacgccc catccagcgc tgacgaaagc aaaacaccag 3540
 cagcagttt tccagttccg tttatccggg caaacatcg aagtgaccag cgaataccctg 3600
 5 ttccgtcata gcgataaacga gtcctgcac tggatggtgg cgctggatgg taagccgtg 3660
 gcaaggcgtg aagtgcctct ggatgtcgct ccacaaggta aacagttgat tgaactgcct 3720
 gaactaccgc agccggagag cgcgggcaa ctctggctca cagtacgcgt agtgcaaccg 3780
 aacgcgacgg catggtcaga agccggcgc acatcgccct ggcagcagtg gctgtggcg 3840
 gaaaacctca gtgtgacgtc cccggcccg tcccacgcca tcccgcacatc gaccaccagg 3900
 10 gaaaatggatt tttgcatacg gctggtaat aagcgttggc aatttaaccg ccagtcaaggc 3960
 ttctttcac agatgtggat tggcgataaa aaacaactgc tgacggcgt gcgcgatcag 4020
 ttcacccgtg caccgttggta acgcacatt ggcgttaagtg aagcgaccgg cattgaccct 4080
 aacgcctggg tcgaacgcgtg gaaggcggcg ggcattacc aggcgcgaagc agcgttgtt 4140
 cagtgacacg cagataact tgctgtatcg gtctgatcg tgcgcgtca cgcgtggcag 4200
 15 catcaggggaa aaacattatt tatacgccgg aaaacatccggg gatgtatgg tagtggtcaa 4260
 atggcgatta ccgttgatgt tgaagtggcg aagcatacac cgcacccggc gccgatgtgc 4320
 ctgaactgcc agctggcgca ggtacgagag cgggtaaact ggctcggtt aaggccgcaa 4380
 gaaaactatac ccgacccgcct tactggccgc tttttgacc gctggatct gccattgtca 4440
 gacatgtata ccccgatcg tttcccgagc gaaaacggtc tgcgcgtcgg gacgcgcgaa 4500
 20 ttgaattatg gcccacacca gtggcgcggc gacttccagt tcaacatcg cgcctacagt 4560
 caacagcaac tgatggaaac cagccatcg catctgtgc acgoggaaaga aggacatagg 4620
 ctgaatatacg acgggttcca tatgggatt ggtggcgacg actcctggag cccgtcaga 4680
 tcggcggaat tccagcttag taccattacc agtggctcg gtgtcaaaaa 4740
 taataataac cgggcagggg gatctttgtt gaaggaacct tacttctgtt gtgtgacata 4800
 25 attggacaaa ctacccatcg attataaag ctctaaggta aatataaaat ttttaagtgt 4860
 ataatgtgtt aaactactga ttcttaattgt ttgtgtatTT tagattccaa cctatggac 4920
 tgatgaatgg gaggcgttggt ggaatgcacat atccagacat gataagatac attgtatgt 4980
 ttggacaaac cacaactaga atgcagtggaa aaaaatgtttt tatttgcggaa atttgcgtt 5040
 ctatgtctt atttgcgttgg attataaact gcaataaaaca agttaacaaac aacaatttgc 5100
 30 ttcattttat gtttcagggt ttttgcgtt gattatgtc tgcggccgca gggcctcgat atacgcctat 5220
 tctacaaatg tggatggct ttttgcgtt gataataatgg ttttgcgtt gtcagggtgc acttttgcgg 5280
 gaaatgtgcg cggacccctc atttgcgtt ttttctaaat acattccaaat atgtatccgc 5340
 tcatgagaca ataaccctga ttttgcgtt gataataatgg aaaaaggaaag agtatacgata 5400
 35 ttcaacattt ccgtgtcgcc aacgcgtggc gttacatcg actggatctc ttattccct ttttgcgtt atttgcctt cctgtttttg 5460
 ctcacccaga aacgcgtggc gttacatcg actggatctc ttattccct ttttgcgtt atttgcctt cctgtttttg 5520
 gtttccaat gatgagactc aacgcggggca agagcaactc ttattccct ttttgcgtt atttgcctt cctgtttttg 5580
 40 actcaccagt cacagaaaag ctgtccataac catgagtgat aacactgggg ccaacttact ttattccct ttttgcgtt atttgcctt cctgtttttg 5640
 cgaaggagct aaccgtttt ttgcacaaca tggggatca tgtaactcgat cttgtatcg 5880
 gggaaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtatcg 5940
 caatggcaac aacgttgcgc aaaaattttt ctggcgtact acttactcta gcttcccgcc 6000
 45 aacaatttaat agactggatg gaggcggata aagttgcgg accacttctg cgctcgccccc 6060
 ttccggctgg ctgggtttt gctgataat ctggagccgg tgagctggg tctcgccgtt 6120
 tcattgcgc actggggcca gatggtaatc cctcccgat cgtatgtt tacacgcacgg 6180
 ggagtccaggc aactatggat gaacgaaata gacagatcg tgagataggt gcctcaactga 6240
 ttaagcatgg gtaactgtca gaccaaggat ttctatataat acttttagatt gattttaaac 6300
 50 ttcatttta attttaaagg atcttaggtt agatcccttt tgataatctc atgaccaaaa 6360
 tcccttaacg tgagtttgc ttcacttgat cgtcagaccc ctagaaaaag atcaaaggat 6420
 cttcttgaga tcctttttt ctgcgcgtt tctgtgttgc gcaaaacaaaa aaaccaccgc 6480
 taccagcggt ggttttttgc cggatcaag agtaccaac tcttttccg aaggtactg 6540
 gtttcagcag aacgcgtggata ccaataactg tccttctatgt ttagccgtt ttagggccacc 6600
 55 acttcaagaa ctctgtatcg ccgcctacat acctcgctt gctaattcctg ttaccagtgg 6660
 ctgctcccgat tggcgataag tcgtgttgc cgggttggc ctcaagacga tagttaccgg 6720
 ataaggcgcg cgggtcgccc tgaacgggg gttcgtgcac acagccacg ttggagcggaa 6780
 cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc acgcttcccg 6840
 aaggggagaaa ggcgcacagg tatccggtaa ggcgcagggt cggAACAGGA gaggcgcacga 6900
 60 gggagcttcc agggggaaac gcttgcgttcc ttatagtc tgctgggttt cgccacccct 6960
 gacttgcgcg tcgattttttgc tgatgcgtt cggggggcg gacccatgg aaaaacgcac 7020
 gcaacgcggc cttttacgg ttccgtggcc ttttgcgttcc ttatgcac atgttcttc 7080
 ctgcgttatac ccctgatttc tggtgataacc gtattaccgc ttgtgtatcg tttgtatcg 7140
 ctgcgcgcag cccgaacgcacc gaggcgcacg agtcaatcg cggaggacgc gaggagcggc 7200
 caatacgcac accgccttc cccgcgcgtt ggccgatcg ttaatgcacg tggcacgaca 7260
 65 gtttcccgat ctggaaagcg ggcgttgatc gcaacgcacat taatgtatcg tagctcactc 7320
 attaggcacc ccaggcttgc cactttatgc ttccggctcg tatgttgcgtt ggaattgtga 7380
 ggggataaca atttcacaca ggaaacagat atgaccatcg ttacgcacaa ctggcgcc 7438

5 <210> 6
<211> 45
<212> DNA
<213> Artificial Sequence

10 <220>
<223> Description of Artificial Sequence: primer C31-1

10 <400> 6
ataagaatgc ggccgccccga tatgacacaaa ggggttgtga ccggg 45

15 <210> 7
<211> 40
<212> DNA
<213> Artificial Sequence

20 <220>
<223> Description of Artificial Sequence: primer C31-3

20 <400> 7
ataagaatgc ggccgcatcc gccgcgtacgt cttccgtgcc 40
25

30 <210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

35 <220>
<223> Description of Artificial Sequence: primer C31-8

35 <400> 8
cccggttggca ggaagcactt ccgg 24

40 <210> 9
<211> 55
<212> DNA
<213> Artificial Sequence

45 <220>
<223> Description of Artificial Sequence: primer C31-9

45 <400> 9
ggatcctcga gccgcgggcg gccgcctacg ccgcgtacgtc ttccgtgccg tccctg 55
50

55 <210> 10
<211> 5711
<212> DNA
<213> Artificial Sequence

55 <220>
<223> Description of Artificial Sequence: vector
pCMV-C31-Int (wt)

60 <400> 10
aaacagtcgg atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
tagtaatcaa ttacggggtc attagttcat agccatata tggagttccg cgttacataa 120
cttacggtaa atggcccgcc tggctgaccg cccaaacgacc cccgcccatt gacgtcaata 180
atgacgtatg ttcccatagt aacgccaata gggatttcc attgacgtca atgggtggac 240
65 tatttacggt aaactgccc a cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcaccaatgcatgaccc 360
tgggacttcc tcaatggca gtacatctac gtattagtc tcgctattac catgggtatg 420

5	cggttttggc agtacatcaa	tgggcgtgga tagcggttg actcacgggg atttccaagt	480
	ctccacccca ttgacgtcaa	tgggagtttgc accatggcacc aaaatcaacg ggactttcca	540
	aaatgtcgta acaactccgc	cccattgacg caaatggcg gttaggcgtgt acgggtggag	600
	gtctataaa gcagagctct	ctggctaact agagaacccaa ctgcttactg gcttatcgaa	660
10	attaatacga ctcactatag	ggagacccaa gctgactcta gacttaatta agcgttggg	720
	tgagtactcc ctctcaaaaag	cgggcatgac ttetgcgcta agattgtcag titccaaaaa	780
	cgaggaggat ttgatattca	cctggccgc ggtgatgcct ttgagggtgg cgcgtccat	840
	ctggtcagaa aagacaatct	ttttgttgc aagcttggg tggcggaggc ttgagatctg	900
15	gccatacact tgagtgacat	tgacatccac tttgccttc tctccacagg tgcactcc	960
	cagggcggcc gcccataatg	acacaagggg ttgtgaccgg ggtggacacg tacgcgggtg	1020
	cttacgacgg tcagtcgccc	gagcgcgaga attcgagcgc agcaagccaa gcgcacacagc	1080
	gtagcgccaa cgaagacaag	gcccgcgacc ttcaagcgcga atgcgagcgc gacggggggc	1140
	gttgcagggt cgtccggcat	ttcaagcgaag cgcgggcac gtcggcggtc ggcacggcgg	1200
20	agcgcggcga gttcaacgc	atccctgaacg aatgcgcgcg cggcggcgc aacatgatca	1260
	ttgtctatga cgttcgcgc	tttcgcgc tgaaggcat ggcgcgatt ccgattgtct	1320
	cggaattgtt cgcctgggc	gtgacgattt ttttcaactca ggaaggcgtc ttccggcagg	1380
	aaaacgtcat ggacctgatt	cacctgatta tgcggctcga cgcgtcgcac aaagaatctt	1440
	cgctgaagtc ggcgaagatt	ctcgacacga agaacctca ggcgcgaaattt ggcgggtacg	1500
25	tcggcggggaa ggcgccttac	ggcttcgagc ttgtttcgga gacgaaggag atcacgcgc	1560
	acggccgaat ggtcaatgtc	gtcatcaaca agcttgcga ctcgaccact ccccttacgg	1620
	gacccttcga gttcggcccc	gacgtaatcc ggtgggtgt ggcgtgagatc aagacgcaca	1680
	aacaccttcc cttaaagccg	ggcagtcgaag ccgcattca cccgggcagc atcacggggc	1740
	ttttaaagcg catggacgct	gacgcgtgc cgacccgggg cgagacgattt gggagaaga	1800
30	ccgcttcaag cgcctgggac	ccggcaacccg ttatgcgaat ctttcggac cgcgttattt	1860
	cgggccttcg cgcgtgggt	atctacaaga agaagccgga cggcacgc accacgaaga	1920
	ttgagggtta cgcatttcg	cgcgacccga tcacgctccg gccggtcgag cttgatttgc	1980
	gaccgatcat cgagcccgct	gagtggatgt agcttcaggc gtgggtggac ggcagggggc	2040
	gcccccaagg gcttcccg	gggcgaaggca ttctgtccgc catggacaag ctgtacttgc	2100
35	agtgtggcgc cgtcatgact	tcgaagcgcg gggaaagatc gatcaaggac tcttaccgt	2160
	gccgtcgccg gaaggtggc	gaccctgtccg cacttggca gcacgaaggc acgtgcac	2220
	tcagcatgac ggcactcgac	aagttcgtt cggaaacgcat ttcaaaag atcaggcaca	2280
	ccgaaggcga cgaagagacg	ttggcgcttc tggggaaagc cgcccgacgc ttccggcaac	2340
	tcactgagc gcctggagaag	agcggcgaac gggcaaccc ttttgcggag cgcggcaca	2400
40	ccctgaacgc ccttgaagag	ctgtacgaag accgcgcggc aggccgtac gacggaccgg	2460
	ttggcagggaa gcacttccgg	aagcaacagg cagcgtac gctccggcag caaggggcgg	2520
	aagagcggct tgccgaactt	gaagccggc aagccccgaa gtttccctt gaccaatgtt	2580
	tcccgaaaga cgccgacgct	gaccggacccg gccttaagtc gtgggggggg cgcgcgtcag	2640
	tagacgacaa ggcgtgttc	gtcggtctct tcttagacaa gatcggtgtc acgaagtcga	2700
45	ctacgggcag gggcagggga	acgcccattcg agaagcgcgc ttctgatcactg tggggaaagc	2760
	cgccgacccg cgacgacgaa	gacgcacccc aggacggcac ggaagacgtt ggcgcgttag	2820
	cgccgccccg gtcgagatc	caggcgcgga tcaataaaag atcattattt tcaatagatc	2880
	tgtgtgttgg tttttgtgt	gccttgggggg agggggaggc cagaatgagg cgcggccaa	2940
	ggggagggggg aggcacaaat	gaccttgggggg gaggggggagg ccagaatgac cttggggggag	3000
50	ggggaggccca gaatggggcg	cgccccccggg taccgagctc gaatttactg ggcgtcggtt	3060
	tacaacgtcg tgacttggaa	aaccctggcg ttaccaact taatgcctt gacgcacatc	3120
	cccccttcgc cagctggcg	aatagcgaag aggccccgcac cgatgcctt tcccaacagt	3180
	tgcgcagctc gaatggcgaa	tggcgctga tgggttattt ttccttacg catctgtgc	3240
	gtatttcaca cgcataatgg	tgcactctca gtacaatctg ctctgtatcc gcatatgtt	3300
55	gccagccccc acacccgcac	acacccgtcg acgcgcctcg acgggttgc ctgtccccc	3360
	catccgttca cagacaagct	gtgaccgtct cggggagctg catgtgtcag aggttttcac	3420
	cgtcatcacc gaaacgcgcg	agacgaaagg gcctcgtat acgccttattt ttataggta	3480
	atgtcatgat aataatggtt	tcttagacgt caggtggcac ttttggggaa aatgtgcgcg	3540
	gaacccctat ttgttttattt	ttctaaatac attcaaaatattt gtatccgtc atgagacaat	3600
60	aaccctgtata aatgtttca	taatatttggaa aaaggaagag tatgagtattt caacatccc	3660
	gtgtcgccct tattcccttt	tttgcggcat tttgccttcc tttttttgtt caccctggaaa	3720
	cgctgggtaa agtaaaagat	gtctgaagatc agttgggtgc acgagtgggt tacatcgac	3780
	tggatctcaa cagcggtta	atcccttgaga gtttgcggcc cgaagaacgtt ttccaaatgt	3840
65	tgagcacttt taaagttctg	ctatgtggcg cggattttac cctgtatttgc gccgttggc	3900
	agcaactcgg tcgcgcata	cactatttctc agaatgactt gtttgcgtac taccaggta	3960
	cagaaaagca tcttacggat	ggcgtacacg taagagaattt atgcgtgtc gccataacca	4020
	tgagtgtataa cacttcggcc	aacttacttc tgacaacgtat cggaggaccg aaggagctaa	4080
	ccgctttttt gcacaaacatg	ggggatcatg taactcgctc tgatgttgg gaaacccggagc	4140
	tgaatgaagc cataccaaac	gacgagcggt acaccacgtt gcctgtatca atggcaacaa	4200
	cgttgcgcaa actattaaact	ggcgaactac ttactctgc ttccggccaa caattaaatag	4260
	actggatgaa ggcgataaaa	tttgcggagc cacttctgcg ctcggccctt cccgttggc	4320
	ggtttatttgc tgataaaatct	ggagccgggtg aggttgggtc tccgggtatc attgcagcac	4380
	tgggccaga tggtaaagccc	tcccgatcg tagttatcta cacgacgggg agtcaggcaa	4440

5 ctatggatga acgaaataga cagatcgctg agatagggtgc ctcactgatt aagcatttgt 4500
 aactgtcaga ccaagtttac tcataatatac ttttagattga tttaaaactt catttttaat 4560
 ttaaaaaggat ctaggtaag atcccttttgc ataatctcat gaccaaaaatc ccttaacgtg 4620
 agttttcggtt ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagato 4680
 10 cttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcgggtgg 4740
 ttgttgcgg ggatcaagag ctaccaactc ttttccgaa ggttaactggc ttccagcagag 4800
 cgcagatacc aaataactgtc ctctctatgt agccgttagt aggccaccac ttcaagaact 4860
 ctgttagcacc gcctacatac ctgcgtctgc taatcctgtt accagtggtc gtcggccatg 4920
 15 gcgataagtgc gtgtcttacc ggggtggact caagacgata gttaccggat aaggcgcagc 4980
 ggtcgggctg aacgggggggt tcgtgcacac agcccgatc ggagcgaacg acctacacccg 5040
 aactgagata cctacagcgt gagctatgag aaagcgccac gcttccgaa gggagaaaagg 5100
 cggacaggtt tccgtaagc ggcagggtcg gaacaggaga ggcacacgagg gagcttccag 5160
 ggggaaacgc ctggtatctt tatagtoctg tcgggtttcg ccacctctga ttgcgcgtc 5220
 gatttttgcg atgcgtcgtca gggggggcga gctatggaa aaacgcgcgaa aacgcggcct 5280
 20 tttacgggtt cctggcctt tgctggcctt ttgcgtcacat gttcttccct gctttatccc 5340
 ctgatttctgt ggataaccgtt attaccgcct ttgagtggc tgataccgcg cggccgcagcc 5400
 gaacgaccga ggcacgcgat tcagtggc aggaagcgga agagcgcaca atacgcaaac 5460
 cgcctctccc cgcgcgttgg cgcatttattt aatgcgtcgtc gcacgcacagg ttcccgact 5520
 25 ggaaagcggg cagtggcgc aacgcattt aatgtgagttt gctcactatc taggcacccc 5580
 aggcttaca ctttatgtttt cccgctcgta tttttgttgg aattgtgagc ggataacaat 5640
 ttcacacagg aaacagctat gaccatgatt acgccaagct agccgggct agcttgcatt 5700
 cctgcagggtt t 5711

30 <210> 11
 <211> 69
 <212> DNA
 <213> Artificial Sequence
 35 <220>
 <223> Description of Artificial Sequence: primer C31-2-2
 <400> 11
 tagaattccg ctgcgagatc taaaccttcc tcttcttctt aggcgcgcgt acgtcttccg 60
 tgccgtcct 69

40 <210> 12
 <211> 5723
 <212> DNA
 <213> Artificial Sequence
 45 <220>
 <223> Description of Artificial Sequence: vector
 pCMV-C31-Lnt(CNLS)
 <400> 12
 cctgcagggtt taaacagttcc gatgtacggg ccagatatac gcgttgcacat tgattattga 60
 50 ctagtttata atagaatca attacgggtt cattatgtca tagcccatat atggagttcc 120
 gcgttacata acttacggta aatggccgc ctggctgacc gccaacgc ccccgccat 180
 tgacgtcaat aatgacgtat ttcccatag taacgcattt agggacattt cattgacgtc 240
 aatgggtgaa ctatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc 300
 caagtacgcc ccctatttgcgtcaatgacg gtaaatggcc cgcctggcat tatgcggcgt 360
 acatgacctt atgggacttt ctacttggc agtacatcta cgtatttagtc atcgcttattt 420
 55 ccatgggtat gcgggtttgg cagtagatca atggggcgtgg atagcggttt gactcacggg 480
 gatttccaaat tctccacccc attgacgtca atgggagttt gtttggcac caaaatcaac 540
 gggactttcc aaaatgtcgt aacaactccg cccatttgcac gcaaatgggc ggtaggcgtg 600
 tacgggtggg ggtcttatata agcagatc tctggcttaac tagagaaccc actgcttact 660
 60 ggcttatacga aattaatacg actcactata gggagaccca agctgactct agacttaattt 720
 aaggcgttggg gtgagttactc ctctcaaaa gcccgcgtca cttctgcgt aagattgtca 780
 gtttccaaaaa acgaggagga tttgatattc acctggcccg cgggtatgcc ttgggggtt 840
 gcccgcgtca tctggtcaga aaagacaatc ttttgggttcaagcttgcgt gttggcagg 900
 ctttgagatct ggccatatac ttgagttaca ttgacatcca cttctgcctt ctctccacag 960
 65 gtgttcactc ccaggccggc cggccgtatc gacacaagggtt gttgtggacc ggggtggacac 1020
 gtacggggtt gcttacgcacc gtcagtgcgt cggacgcgcg aattcgacgc cagcaagccc 1080
 agcgcacacag cgtacgcaca acgaagacaa ggccggccgac ttccagcgcg aagtcgacgc 1140
 cgacgggggc cgggtcagggt tttcagcgcgaa gcccggggca cgtcggcgtt 1200

	cgggacggcg gaggccccgg	agttcgaacg catctgaac	gaatcccgcg cccggcggtc	1260
	caacatgatc attgtctatg	acgtgtcgcg ctttctcgcc	ctgaaggtaa tggacgcgat	1320
5	tccgattgtc tcggaattgc	tcgccttggg cgtgacgatt	gtttccactc aggaaggcgt	1380
	cttccggcag gaaaacgtca	tggacctgat tcacctgatt	atgcggctcg acgcgtcgca	1440
	caaagaatct tcgctgaagt	cggcgaagat tctcgacacg	aagaaccttc agcgcgaatt	1500
	gggcgggtac gtcggcggga	aggcgctta cggcttcgag	cttgtttcgg agacgaaga	1560
	gatcacgcgc aacggccgaa	tgttcaatgt cgtcatcaac	aagcttgcgc actcgaccac	1620
10	tccccttacc ggacccttcg	agttcgagcc cgacgtaatc	cggtgggtgt ggcgtgagat	1680
	caagacgcac aaacaccttc	ccttcaagcc gggcagtcaa	gcogccattc acccgggcag	1740
	catcacgggg ctttgtaagc	gcatggacgc tgacgcccgt	ccgaccgggg gcgagacgt	1800
	tgggaagaag accgcttcaa	gccccttggg cccggcaacc	gttatgcgaa tccttcggga	1860
	cccgcttatt gccccgttcg	ccgtgaggt gatctacaag	aagaagccgg acggcacgccc	1920
15	gaccacgaag attgagggtt	accgattca cgcgcacccg	atcacgttcc gcccgggtcga	1980
	gcttgatttc ggacgatca	tcgagccgc tgagtgttat	gagcttcaagg cgtgggttga	2040
	ccgcaggggg cgcgcacagg	ggcttcccg ggggcaagcc	attctgtccg ccatggacaa	2100
	gctgtacttc gagtgtggcg	ccgtcatgac ttcaagcgc	gggaaagaat cgatcaaga	2160
20	ctcttaccgc tgccgtcgcc	ggaagggtgt cgacccgtcc	gcaccttggc agcacgaagg	2220
	cacgtcaac gtcagcatgg	ccggactcga caagttcggt	gccaacgcgca tcttcaacaa	2280
	gatcaggcac gccgaaggcg	acgaagagac gttggcgctt	ctgtggaaag ccgcccgcacg	2340
25	cttcggcaag ctcactgagg	cgcttgagaa gagcggcgaa	cgggcgaccc ttgttgcgga	2400
	gcgcgcccac gcccgtaaac	cccttgaaga gctgtacgaa	gaccgcgcgg caggcgcgt	2460
	cgacggaccc gttggcagga	agcacttccg gaagcaacag	gcagcgctga cgctccggca	2520
	gcaaggggcg gaagagcggc	ttgccgaact tgaagccgcc	gaagccccga agttccccct	2580
	tgaccaatgg ttccccgaag	acgcccacgc tgacccgacc	ggccctaagt cgtgggtggg	2640
30	gcgcgcgtca gtagacgaca	gggggcaggg aacgcccata	agcgcgtt cgtggggctc ttctgtagaca	2700
	cacgaagtgc actacggggca	acgacgacga agacgacgca	gagaagcgcg cttcgatcac	2760
	gtgggcgaag cccgcgaccg	agaagggtta gactctcgag	caggacggca cggaaagacgt	2820
	agccccgcct aagaagaaga	atctgtgtg tggttttttt	atccaggcgc ggatcaataa	2880
35	aagatcatta ttttcaatag	aagggggagg gggaggccag	tgtgccttgg gggagggggg	2940
	ggccagaatg aggccgggca	gagggggagg ccagaatgag	aatgacccctt ggggaggggg	3000
	aggccagaat gaccttgggg	ttttacaacg tcgtgactgg	gcccgcaccc ggttacccgag	3060
	ctcgaattca ctggccgtcg	atccccctt cgcagctgg	aaaaaccctg gcgttaccca	3120
	acttaatcgc cttgcagcac	agttgcgcag cctgaatggc	cgtaatagcg aagaggcccg	3180
40	caccgatcgc cttttccaaac	gcggtatattc acaccgcata	gaatggcgcc tgatgcggt	3240
	ttttctcctt acgcatctgt	taagccagcc cgcacacccg	tggtgcactc tcagtacaat	3300
	ctgtctgtat gccgcatagt	ccggatccgc ttacagacaa	ccaacacccg ctgacgcgc	3360
	ctgacgggt tgcgtctcc	ccacgtcattt accgaaacgc	gctgtgaccg tctccggag	3420
45	ctgcgtgtt cagaggttt	ttaatgtcat gataataatg	gcgagacgaa agggcctcg	3480
	gatacgccctt ttttatagg	gcccgaaccc tatttgttta	ttttctttaga cgtcaggtgg	3540
	cacttttcgg gaaatgtgc	aataaccctg ataaatgcctt	ttttctaaa tacattcaaa	3600
	tatgtatccg ctcatgagac	tccgtgtcgc ctttattcc	caataatatt gaaaaaggaa	3660
	gagtatgagt attcaacatt	aaacgcgtgt gaaagtaaaa	tttttgcgg cattttgcct	3720
	ttctgtttt gtcaccccg	actggatct caacagcggt	gatgtcaag atcagttggg	3780
50	tgacacgtt ggttacatcg	tgatgagac ttttaagggt	aagatccctt agagtttgc	3840
	ccccgaagaa cgttttccaa	aagagcaact cggcgccgc	ctgtctgtg ggcgggtatt	3900
	atcccgtatt gacccgggc	tcacagaaaaa gcatcttacg	atacactatt ctcagaatga	3960
	cttgggttag tactcaccag	ccatgagtgta taacactgcg	gatggcatga cagtaagaga	4020
55	attatgcagt gtcgcataa	taaccgcctt tttgcacaaac	gccaacttac ttctgacaaac	4080
	gatcgaggaa ccgaaggagc	agctgaatga agccatacca	atggggatc atgttaactcg	4140
	ccttgatctt tgggaaccgg	caacgttgcg caaactattt	aacgacgacg gtgacaccac	4200
	gatccctgtt gcaatggca	tagactggat ggaggcggt	actggcgaac tacttactt	4260
	agcttcccgg caacaattaa	gctggtttat tgctgataaa	aaagttgcag gaccacttct	4320
	gctctcgccg cttccggctg	cactggggcc agatggtaag	tctggagccg gtgagcgtgg	4380
60	gtctcgccgt atcattgcag	caactatggta tgaacgaaat	ccctcccgta tcgtagttat	4440
	ctacacgcg gggagtcagg	ggttaactgtc agaccaagtt	agacagatcg ctgagatagg	4500
	tgcctactg attaaggatt	taattaaaaa gatcttaggt	tactcatata tacttttagat	4560
	tgataaaaa cttcattttt	catgtacccaa atcccttaac	aagatccctt ttgataatct	4620
	catgaccaaa atcccttaac	gtgagtttc gttccactga	gctcagacc ccgtagaaaa	4680
	gatcaaaggaa ttttcttgag	atcccttttt tctgcgtcgta	atctgtctgt tgcaaaacaaa	4740
65	aaaaccaccc taccagcgg	tggtttttt gccgatcaa	gagctaccaa ctctttttcc	4800
	gaaggtaact ggcttcgca	gagcgcagat accaaaatact	gtccttctag ttagccgt	4860
	gttaggcccac cacttcaaga	actctgtac accgcctaca	tacctgtctc tgctaatcc	4920
	gttaccatgt gctgtcgca	gtggcgataa gtcgtgtt	accgggttgg actcaagacg	4980
	atagttaccg gataaggcgc	agcggtcggg ctgaacgggg	ggttcgtcga cacagccccag	5040
	cttggagcga acgacctaca	ccgaactgag atacctacag	cgtgagctat gagaaagcgc	5100
	cacgcctccc gaaggagaaa	aggcggacag gtatccggta	agcggcagggtt cggaaacagg	5160
	agagcgcacg agggagttc	cagggggaaa cgcttggat	ctttatagtc ctgtcggtt	5220

tcgccaccc tgacttgagc gtcgatttt gtatgtctcg tcaggggggc ggagcctatg 5280
 gaaaaacgcc agcaacgcgg ccttttacg gttctggcc ttttgcgtgc cttttgcctca 5340
 5 catgttcttt cctgcgttat cccctgatcc tggataac cgtattaccg cttttgagtg 5400
 agctgatacc gctcgccgca gccgaacgac cgagcgcagc gagtcaatgca gcgaggaagc 5460
 gaaagagcgc ccaatacgca aaccgcctct ccccgccgt tggccgattc attaatgcag 5520
 ctggcacgac aggttcccg actggaaagc gggcgtgag cgcaacgcaa ttaatgttag 5580
 ttatctact cattaggcac cccaggctt acactttatg cttccgctc gtatgttg 5640
 tggaaattgtg agcgataac aatttcacac agggaaacagc tatgaccatg attacgcca 5700
 10 gtagcccg gctagcttgc atg 5723

10

<210> 13
 <211> 4960

15 <212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
 20 pCMV-Cre

<400> 13

aaacagtcgg atgtacgggc cagatatacg ctttgacatt gattattgac tagttattaa 60
 tagtaatcaa ttacggggtc attagttcat agccatata tggagttccg ctttacataa 120
 25 cttagcgta atggcccgcc tggctgaccg cccaaacgacc cccgcccatt gacgtcaata 180
 atgacgtatg ttcccatagt aacgccta aggacttcc attgacgtca atgggtggac 240
 tatttacggt aaactgccta ctggcagta catcaagtgt atcatatgcc aagtacgccc 300
 cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgaccta 360
 tgggactttc ctacttggca gtacatctac gtattatgc tcgcttattac catggtgatg 420
 cggttttggc agtacatcaa tggcgtgga tagcgggttg actcacgggg atttccaagt 480
 30 ctccacccca ttgacgtcaa tggagtttgc tttggcacc aaaatcaacg ggactttcca 540
 aatgtcgta acaactccgc cccattgacg caaatggcgc gtaggcgtgt acgggtggag 600
 gtctatataa gcagagctt ctggctact agagaaccca ctgcttactg gcttatcgaa 660
 attaatacga ctcaactatag ggagacccaa gctgactcta gacttaatta agcgttgggg 720
 35 ttagtactcc ctctcaaaag cgggcatgac ttctgcgcta agattgtcag ttccaaaaaa 780
 cgaggaggat ttgatattca ctggcccgcc ggtgatgcct ttgagggtgg cccggtccat 840
 ctggtcagaa aagacaatct ttttgttgc aagcttgagg tggcaggc ttgagatctg 900
 gcacatacact tgagtgcacat tgacatccac ttgccttc tctccacagg tgcacttcc 960
 cagggcggcc tcgaccatgc ccaagaagaa gaggaaagggtg tccaaattac tgaccgtaca 1020
 40 cccaaatttt cctgcattac cggtcgtatgc aacgatgtat gaggttcgca agaacctgtat 1080
 ggacatgttc agggatgcgc aggcttttc tgacatacc tggaaaatgc ttctgtccgt 1140
 ttggcgttc tggcggcat ggtgcaagttt gaataacccgg aatgttttcc cccgagaacc 1200
 tgaagatgtt cgcgattatc ttctataatct tcaggcgcgc ggtctgcag taaaaactat 1260
 45 ccaccaacat ttggcccagc taaacatgt tcatcgatgg tccgggtgc cacgaccaag 1320
 tgacagcaat gctgtttcac tggttatgcg gcgatccga aaagaaaacg ttgatgccgg 1380
 tgaacgtgoa aaacaggctc tagcgttgc acgactgtat ttcgaccagg ttgttcaact 1440
 catggaaaat agcgatcgct gccaggatatacgtatctg gcatattctgg ggattgctta 1500
 taacaccctg ttacgtatag ccgaaattgc cagatcagg gttaaagata tctcacgtac 1560
 50 tgacgggtgg agaatgtttaa tccatattgg cagaacgaaa acgctgttta gcacccgagg 1620
 ttagagaaat gcacttagcc tgggggttaac taaactggc gacgatggta tttccgtctc 1680
 tggtagatgt gatgtccga ataactacct gtttgcgg gtcagaaaaa atgggtttgc 1740
 cgcgcctatc gccaccagcc agctatcaac tcgcgcctg gaaggatgtt ttgaagcaac 1800
 55 tcatcgattt atttacggcg ctaaggatga ctctggatcg agataactgg cctggctctgg 1860
 acacagtgcg cgtgtcgagg ccgcgcgaga tatggcccg gctggagttt caataccgga 1920
 gatcatgca gctgggtggc ggaccaatgt aaatattgtc atgaactata tccgtaacat 1980
 ggatagtgaa acaggggcaaa tggtgcgcct gctggaagat ggcgattagc cattaacgcg 2040
 taaatgattt cagatccact agttcttaggg cccgcgtcgac ctcgagatcc aggcgcggat 2100
 60 caataaaaaga tcattttt caatagatct gtgtgttgc ttttgcgtgc ctttggggga 2160
 gggggaggcc agaatgaggc gggcccaagg gggaggggga ggcggaaatg accttggggg 2220
 agggggaggcc cagaatgacc tggggggagg gggaggccag aatgaggcgc gccccccggg 2280
 accgagctcg aatttactgg ccgtcgatcc acaacgtcg gactggggaa accctggcgt 2340
 taccctaaatc aatgccttgc cagcacatcc cccttcgc gctggcgta atagcgaaga 2400
 gccccgcacc gatgcgcctt cccaaacagtt ggcgcgcctg aatggcgaat ggcgcctgat 2460
 gcggtatccc ctccctacgc atctgtgcgg tatttcacac cgcataatggt gcaactctca 2520
 tacaatctgc tctgtgcgc catagttaa ccagccccga caccgcctaa caccgcgtca 2580
 65 cgcgccttgc cgggttgc tgcctccggc atccgccttac agacaagctg tgaccgtctc 2640
 cgggagctgc atgtgtcaga gttttcacc gtcacatcccg aaacgcgcga gacgaaagg 2700
 ctcgtatacgccatcc tataaggtaa tgcgtatgata ataatggttt ctttagacgtc 2760

5	agggtggcact tttcgaaaa atgtgcgcgg aacccttattt tgtttatttt tctaaataca 2820
	ttcaaataatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 2880
	aaggaagagt atgagtattc aacatttccg tgcggccctt attcccttt ttgcggcatt 2940
	ttgccttcct gttttgttc acccagaaac gctgggtaaa gtaaaagatg ctgaagatca 3000
	gttgggtca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 3060
	tttcgcgccc gaagaacgtt ttccaatgtt gggactttt aaagttctgc tatgtggcgc 3120
	ggtattatcc cgtattgtac ccgggcaaga gcaactcggt cgccgcatac actattctca 3180
	gaatgacttg gttgagttt caccagtac agaaaagcat ttacggatg gcatgacagt 3240
10	aagagaatia tgcaatgtt ccataaccat gggataao actgoggcca acttacttct 3300
	gacaacgatc ggaggaccga aggagctaac cgctttttt cacaacatgg gggatcatgt 3360
	aactcgccctt gatcggttggg aaccggagct gaatgaagcc atacaaacg accagcgtga 3420
	caccacgtt cctgtacaa tggcaacaaac gttgcgc当地 ctatctact gcaactact 3480
	tactctatgtt tccggcaac aattaaataga ctggatggg gggataaaag ttgcaggacc 3540
15	acttctcgcc tccggccctt cggctggctt gtttattgtt gataaaatctg gggccgttga 3600
	ggttgggtctt cgcgtatca ttgcagactt gggccagat ggttggccctt cccgtatctg 3660
	agttatctac acgacgggaa gtcaggcaac tattggatgaa cggaaatagac agatcgctga 3720
	gataggtgcc tcactgatta agcattgtt actgtcagac caagttactt catatataact 3780
	tttagattgtt taaaacttc attttaatt taaaaggatc taggtgaaga tccttttga 3840
20	taatctcatg accaaaatcc cttaacgtt gtttgcgtt cactgagcgtt cagaccctt 3900
	agaaaagatc aaaggatctt ctttgcgtt cgcgtatct gctgcttgc 3960
	aacaaaaaaa ccacccgtac cagcggtt gttttgcgtt gatcaagagc taccatct 4020
	tttccgaag gtaactggct tcagcagagc gcaagatcca aataactgtcc ttcttagtgt 4080
	gccgttagtta gcccaccact tcaagaaactc tggtagcaccg cttacatacc tccgtctgt 4140
25	aatcctgtt ccagtggctt cggccagttt ggttggactt 4200
	aagacgatag ttaccggata gggcgacgat gtcgggctt cgtgcacaca 4260
	gcccgacgtt gggcgacgat cttacaccga actgagataac ctacagcgtt agctatgaga 4320
	aagcgccacg cttccggaaag ggagaaaggc ggacaggatc cggtaagcg gcagggttgc 4380
	aacaggagag cgcacggagg agcttccagg gggaaacgcg tggatctttt atagtcctgt 4440
30	cgggttgcg caccctgttgc ttgagcgtt gttttgttgc tgcgttgc 4500
	cctatggaaa aacggccagca acggccctt tttacgggtt ctggccctt gctggccctt 4560
	tgttcacatg ttcttgcgtt cttatcccc ttgattctgtt gataaccgtt ttaccgcctt 4620
	tgagtgttgcg tataccgttgc cccgcaggcc aacgaccgag cgcagcgtt cgttgc 4680
	ggaagcgaaa gaggcccaa tggcaaaacc gccttcccc ggcgttgc cgattcatta 4740
35	atgcagctgg cacgacaggat tttccgactt gaaagcggtt agtggcgca acgcaattaa 4800
	tgttgggttgcg attgtgagcg gataacaattt tcacacagga aacagctatg accatgatta 4920
	cgccaaagcta gccccggctt gcttgcgttgc ctgcagggtt 4960

40 <210> 14
 <211> 3858
 <212> DNA
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: vector pRK50

40	<400> 14
	aaacagtccg atgtacgggc cagatataacg cgttgcattt gattattgac tagtttattaa 60
50	tagtaatcaa ttacgggtt attagttcat agccatata tggagttccg cgttacataaa 120
	cttacggtaa atggccggcc tggctgaccg cccaaacgacc cccggccatt gacgtcaata 180
	atgacgtatg ttcccatatgtt aacgccaata gggactttcc attgacgtca atgggtggac 240
	tatttacgtt aaactggccca cttggcagta catcaagtgtt atcatatgcc aagtacgccc 300
	cctattgacg tcaatgacgg taaaatggccc gcttggcattt atgcccagta catgaccctt 360
55	tgggacttttcc ctacttggca gtacatcttcat gtattagtca tgcgttattac catgggtatg 420
	cggttttggc agtacatcaa tgggctgttga tagcggtttt actcacgggg atttccaagt 480
	ctccacccca ttgacgttca tgggagttt ttttggcacc aaaatcaacg ggactttcca 540
	aaatgtcgtt acaactccgc cccatttgcg caaatggccg gtggcggtt acgggtggag 600
60	gtcttatataa gcagagctt ctggcttaactt agagaacccca ctgcttactt gcttatcgaa 660
	attaataacg ctcactatag ggagacccaa gctgacttca gacttaattt aacgttgggg 720
	tgagtactcc ctctcaaaaag cggggatgttgc ttctggcgttca agatgttgcg tttccaaaaa 780
	cgaggaggtt ttgtatattca cctggccgc ggtgtatgttgc ttgggggttgc cccgttccat 840
65	ctgggtcagaa aagacaatctt ttttgggttca aagcttgggg tttggggcggc ttgagatctg 900
	gccataactt tgagtgttgc tggatccac tttggcttcc tctccacagg tttccactcc 960
	caggggcgccc gctgtcgaccc tggatccag ggcggatca ataaaagatc attatttca 1020
	atagatctgtt gttgtgggtt ttttgggttca aatggggccg aatggggccg 1080
	ggccaaagggg gaggggggagg ccagaatgac cttggggggag gggggggccca gaatgaccc 1140

10

gggggagggg gaggccagaa tgaggcgcgc ccccggtac cgagctcgaa ttcactggcc 1200
 gtcgtttac aacgtcgtga ctgggaaaac cctggcgta cccaaacttaa tcgccttgca 1260
 gcacatcccc ctttcgcccag ctggcgtaat agcgaagagg cccgcaccga tcgcccottcc 1320
 5 caacagttgc gcagcctgaa tggcgaatgg cgctctatgc ggtattttct ccttacgcac 1380
 ctgtgcggta tttcacaccg catatggtc actctcgta caatctgc tgcgtccgca 1440
 tagttaagcc agccccgaca cccgccaaca cccgctgacg cgcctgacg ggcttgtctg 1500
 ctcccgcat cccgttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg 1560
 ttttcaccgt catcaccgaa acgcgcgaga cgaaggggcc tcgtgatacg cctattttta 1620
 taggttaatg tcatgataat aatggttct tagacgtcag gtggcaactt tcggggaaat 1680
 10 gtgcgcggaa cccctattt tttattttcaaaatacatt caaatatgtt tccgctcatg 1740
 agacaataaac cctgataaaat gottcaataa tattgaaaaaa ggaagagttt gaggatttcaa 1800
 catttcgtg tcgccttat tcccttttgcgcattttt gccttcctgt ttttgctcac 1860
 ccagaaacgc tggtggaaatg aaaagatgtt gaagatcagt tgggtgcacg agtgggttac 1920
 atcgaactgg atctcaacag cggtaagatc ctggatggat ttcgccccga agaacgtttt 1980
 15 ccaatgtatgac gcaacttttta agttctgtt tttggcgccg tatttcccg tatttgcggcc 2040
 gggcaagagc aactcggtgc cccgcatacac tatttctcaga atgacttggt tgagtactca 2100
 ccagtcacag aaaagcatct tacggatggc atgacagatca gagaattatg cagtgcgtcc 2160
 ataaccatga gtgataacac tgcggccaaat ttaacttctga caacgatcg aggaccgaag 2220
 20 gagctaaaccg ctttttgca caacatgggg gatcatgtaa ctcgccttgc tcgttgggaa 2280
 ccggagctga atgaagccat cccaaacgac gagcgtgaca ccacgatgcc tgtagcaatg 2340
 gcaacaacgt tgcgcaactt attaacttgc gaaactactt ctctagatc cccggcaacaa 2400
 ttaatagact ggatggagc ggataaaatg gcaggaccac ttctgcgttc gggcccttccg 2460
 gctggctgtt ttatttgcgtt taaatcttgcg gcccgtgagc gtgggtctcg cggatcatt 2520
 25 gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 2580
 caggcaacta tggatgaacg aaatagacag atcgtcgatc taggtgcctc actgatattag 2640
 cattggtaac tgcgtacca agtttactca tatatactttt agattgattt aaaacttcat 2700
 ttttaatttta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct 2760
 taacgtgagt ttgcgttccg ctgagcgtca gaccccgtag aaaagatcaa aggtatctt 2820
 30 tgagatcctt ttttctgcg cgtaatctgc tgcttgcaaa caaaaaaaaacc accgctacca 2880
 ggggtggttt gtttgcggg tcaagagctt ccaactctttt ttcggaaatg aactggcttc 2940
 agcagacgac agataccaaa tactgtcctt ctatgttagc cgtatgttgg ccaccacttc 3000
 aagaactctg tagcaccggc tcatatcttgc gctctgtcaa tcctgttacc agtggctgt 3060
 gcccgtggc ataagtctgt tcttaccggg ttggactcaa gacgatgtt accggataag 3120
 35 gcgccgggtt cgggctgaaac ggggggttcg tgcacacagc ccagcttgg gcaacgacc 3180
 tacaccgaac tgagataacctt acagcgttag ctatgagaaa gcccacacgt tcccgaagg 3240
 agaaaaggccg acaggtatcc ggtaaaggccg agggtcgaa caggagacg caccggggag 3300
 40 cttccagggg gaaacgcctg gtatctttt agtctgtcg ggtttcgcca cctctgactt 3360
 gagcgtcgat ttttgcgtt ctcgtcaggg gggccggagcc tatggaaaaaa cggccacaa 3420
 gcccgtttt tacggttcc tgccttttgc tggcccttttgc ctcacatgtt cttctctgcg 3480
 ttatcccctg attctgttgc taaccgtatt accgccttttgc agtgcgtga taccgctcgc 3540
 45 cgccaggccaa cggccggcg cagcgtatc gtgagcgttgc aagcggaaaga gcccacac 3600
 cgcggccgc ctctccccgc gcttggccg attcatatgc gcaatgttgc cgcgtggca cgcacagg 3660
 cccgacttgg aaggccggcag tgagcgttgc gcaatgttgc tgatgttagt cactcattag 3720
 gcaccccccgg ctttacactt tatgtttccg gctcgtatgt tttgttgcgtt tttgttgcgtt 3780
 taacaatttc acacaggaaa cagctatgac catgattacg ccaagcttagc cccggcttagc 3840
 ttgcgttgcctt gcagggtt 3858

<210> 15

50 <211> 6257

<212> DNA

<213> Artificial Sequence

<220>

55 <223> Description of Artificial Sequence: vector
pRK64(deltaCre)

<400> 15

cgtcatcacc gaaacgcgcg aggcagctgt ggaatgtgtg tcagtttaggg tttggaaatg 60
 60 ccccaaggctc cccagcaggc agaagtatgc aaagcatgca tctcaatttgc tcagcaacca 120
 ggctccccag caggcagaag tttgtcaatgc atgcatctca attagtctgc aaccatagtc 180
 ccccccctaa ctccgcctt cccgccttactccgccttca gttccgccttca ttctccgc 240
 catggctgac taattttttt tatttgcgtt gaggccggagg cccgccttgc ctaggaacac 300
 65 tegacgacac tgcagagacc tacttctactt acaacccgttca cagttctgtt accagatggg 360
 tgagggtggag tacgcgcggc gggggccca gggccacggcc tggccacccgc accgcggctt 420
 cgagaccgtc acgaatagat ccataacttgc tgcgtatgtt cattatacgtt agttataccg 480
 ggcacccatgtt gtcgtcgactt ggcctgttgc ttacaacgtc gtgacttggg 540

aaaccctggc gttacccaac ttaatgcct tgagcacat cccccttcg ccagctggcg 600
 taatagcgaa gaggccccca ccgatcgccc ttcccaacag ttgcgcagcc tgaatggcga 660
 atggcgctt gcctggttc cggcaccaaga aeggggtccg gaaagctggc tggagtgcga 720
 5 tcttccttag gccgatactg tcgtcgccc ctcaaaactgg cagatgcacg gttacgatgc 780
 , gcccattctac accaacgtaa ctatccat tacggtaat ccggcggtt ttcccacgga 840
 gaatccgacg gttgttact cgctcacatt taatgttgat gaaagctggc tacaggaagg 900
 ccagacgcga attatttttg atggcgtaa ctcggcggtt catctgtgtt gcaacggcg 960
 10 ctgggtcggt tacggccagg acagtcgtt gccgtctgaa ttgacactga ggcgcatttt 1020
 acgcgcgcga gaaaaccgcg tcgcgggtat ggtgtcggt tggagtgcacg gcagttatct 1080
 ggaagatcgat gatatgtggc ggtatgaggcg cattttcgat gacgtctgt tgctgcataa 1140
 accgactaca caaatcagcg atttccatgt tgccactcgc ttaatatgtat atttcagccg 1200
 cgctgtactg gaggctgaag ttcagatgtg cggcgagttt cgtgactacc tacgggtaa 1260
 agtttctta tggcagggtg aaacgcagg cgcacggcg accgcgcctt tcggcggtiga 1320
 15 aattatcgat gaggctgggtg gttatgcgca tcgcgtcaca ctacgtctgaa acgtcgaaaa 1380
 cccgaaactg tggagcgccg aaatcccgaa tcttatactgt ggggtgtt aactgcacac 1440
 cgccgacgac acgctgattt aagcagaagc ctgcgtatgc ggttcccgcg aggtgcggat 1500
 tgaaaatgtt ctgcgtctgc tgaacggcaa gccgttctg attcgaggcg ttaaccgtca 1560
 cgagcatcat cctctgcgtt gtcaggtcat ggtatgaggcg acgtatgtgc aggtatct 1620
 20 gctgtatgaa cagaacaact ttaacgcgtt ggcgtgttgc cattatccga accatccgt 1680
 gtggtacacg ctgtgcgacc gtcacggctt gatgtgttgc gatgaagcca atattgaaa 1740
 ccacggcatg gtgcataatga atcgtctgac cgatgatccg cgctggctac cggcgatgag 1800
 cgaacgcgtt acgcaatgg tgcagcgccg tcgtatcac ccgagtgta tcatctggc 1860
 gctggggat gaatcaggcc acggcgctaa tcacgcacgcg ctgtatcgat ggatcaa 1920
 25 tgcgtatgttcccgccg tgcgtatgttgc aggccggcg ggcgacacca cggccacccga 1980
 tattatttgc cgcgtatgttgc ttcgcgtacc tggagagacg cccttcccg cggttgcgaa 2040
 atggtccatc aaaaatggc atacgccccac gcgatgggtt acagtcgttgcggttcccgctt 2100
 tcagtatccc cgtttacagg ggggtttcgatgggttcccgctt gttttggcg atacgcccga 2160
 30 atatgtatgaa aacggcaacc ttcgtatgttgc ttacggccgtt acggctctgttcccgctt 2220
 cgatcgccat ttctgtatgttgc acggctctgttcccgctt aaatactggc aggctttcg 2280
 gacggaagca aaacaccaggc acgatgtttt ccgtttccgtt ttcgtatgttgc acggccgttcccgctt 2340
 agtgaccaggc gaataacctgtt gtcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2400
 35 gctggatgtt aagccgctgg acagtttgcggtt ccgtttccgtt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2460
 aactaccgcg aactgtctgtt gtcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2520
 agtacgcgtt gtgcacccgcg acggccgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2580
 40 gcaacgttgg cgtctggccg cccgcattcg accaccaggcg atggatccat ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2640
 atttaaccgc cagtcaggctt gacgcccgtt cgcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2700
 gacgcccgtt cgcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2760
 45 agcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2820
 gatgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2880
 gcatccggcg cggatggcc ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 2940
 50 gtcggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3000
 ctgggatctg ccattgttagt gtcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3060
 gcgctgcggg acgcgcgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3120
 55 caacatcagc cgttacagtc aacagcaact gatggaaacc ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3180
 cgccggaaaggc ggcacatggc ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3240
 60 ctccctggagc cgcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3300
 gtttgtctgg tgcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3360
 65 tttaaaaatt ttaatgttgc agattccaaact ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3420
 ggcctcgta tacggctatt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3480
 tcagggtggca ctttccgggg cattcaataatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3540
 aaaaaggaaatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3600
 tttaaaaatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3660
 65 tttaaaaatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3720
 agattccaaact ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3780
 ataagataca ttgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3840
 atttgtgaaa ttgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3900
 gttaaacaaca acaattgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 3960
 tttaaaaatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4020
 ggcctcgta tacggctatt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4080
 60 tcagggtggca ctttccgggg cattcaataatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4140
 aaaaaggaaatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4200
 tttaaaaatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4260
 65 tttaaaaatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4320
 agtttcgccc cccgaaagacg gtttccaatg atggacactt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4380
 gcggtattat cccgtattgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4440
 cagaatgact tgggtgagta ctcaccagtc acagaaaagc atcttacggatccat ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4500
 cagaatgact tgggtgagta ctcaccagtc acagaaaagc atcttacggatccat ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt ttcgtatgttgc ttcgtatgttgc acggatgttcccgctt 4560

5 gtaagagaat tatgcagtgc tgccataaacc atgagtgata acactgcggc caacttactt 4620
 ctgacaacga tcggaggacc gaaggagcta accgcctttt tgcacaacat gggggatcat 4680
 gtaactcgcc ttgatcggtt ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt 4740
 gacaccacga tgcctgttagc aatggcaaca acgttgcgc aactattaac tggcaacta 4800
 cttactctag cttccggca acaattaata gactggatgg aggccgataa agttgcagga 4860
 10 ccacttctgc gctcgccct tccggctggc tggttattt ctgataaattc tggagccgt 4920
 gagcgtgggt ctcgcggat cattgcagca ctggggccag atggtaagcc ctcccgatc 4980
 gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag acagatcgt 5040
 gagataggtg cctcaactgtat taagcatgg taactgttag accaagttt ctcataatata 5100
 15 ctttagatgg attttaaaact tcattttaa tttaaaagga tcttaggtgaa gatcctttt 5160
 gataatctca tgaccaaaat cccttaacgt gagtttctgt tccactgagc gtcagacccc 5220
 gtagaaaaaa tcaaaggatc ttcttgagat ccttttttc tgcgcttaat ctgtctgtt 5280
 caaaaaaaaa aaccacccgt accagcgggt gtttgittgc cggatcaaga gtcaccaact 5340
 20 cttttccga aggttaactgg cttcagcaga ggcagatac caaatactgt cttcttagt 5400
 tagccgtatg tagggcacca cttaaaggaaact cttcagatcata cttcgctctg 5460
 ctaatctgt taccagtggc tgctgcctgt ggccataaact cttcgctatc cgggttggac 5520
 tcaagacgt agttaccggc taaggcgcag cggtcggct gaacgggggg ttctgtgcaca 5580
 cagcccgact tggagcgaac gacctacacc gaactgagat acctacagcg tgagctatga 5640
 25 gaaagcgcga cgcttcccga agggagaaag gcgacaggat atccggtaag cggcagggtc 5700
 ggaacaggag agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtct 5760
 gtcgggtttc gccacctctg acttgagctt cgtgtcttac cgggttggac 5820
 agcctatgga aaaacgcgcag caacgcggcc ttttacggt tccctggcct ttctggcct 5880
 tttgctcaca tggttttcc tgcgttatacc cctgattctg tggataaccg tattaccggc 5940
 30 tttgagtgag ctgataccgc tggccgcagg cgaacgaccg agcgcacgc gtcagtgagc 6000
 gaggaagcgg aagagcgcce aatacgcaaa ccgcctctcc cccgcgttg gccgattcat 6060
 taatgcagct ggcacgacag gtttcccgc tggaaagcgg gcagtgagcg caacgcatt 6120
 aatgtgagtt agctcaactca ttaggcaccc caggcttac actttatgtc tccggctctg 6180
 atgttgtgtg gaattgtgag cggataacaa tttcacacag gaaacagcta tgaccatgt 6240
 tacgccaaga tggcgcg 6257

30

<210> 16

<211> 6252

<212> DNA

35 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pRK64 (deltaInt)

40

<400> 16

cgtcatcacc gaaacgcgcg aggcagctgt ggaatgtgtg tcagtttaggg tggaaagt 60
 cccccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag tcagcaacca 120
 45 ggctccccag caggcagaag tggcaaaacg atgcattctca attagtcgc aaccatagtc 180
 cccgccttaa ctccggccat cccgcctta actccgccta gttccgccta ttctccgc 240
 catggctgac taatttttt tatttgcgca gaggccgagg cccgcctcgcc ctaggaacac 300
 tgcacgacac tgcagagacc tacttcacta acaaccggta cagttctgtt accagatggg 360
 tgaggtggag tacgcgcgg gggagccaa aggttacccc agttggcga ctactccga 420
 50 aaaccgcctc tggatccata acttcgtata gcatacattt tacgaagttt taccggcga 480
 ccatggtcgc gagtagcttgc gcaactggccg tggtttaca acgtcgatc tggaaaacc 540
 ctggcgttac ccaacttaat cgccttcgcg cacaatcccc tttcgccagc tggcgtaata 600
 gcgaaagaggc cgcacccgtt cgcctttcc aacagttgcg cagcctgaat ggcgaatgca 660
 gctttgcctg gttccggca ccagaagcgg tggccggaaag ctggctggag tgcgtatctc 720
 55 ctgaggccga tactgtcgcc tttccctcaaa actggcagat gcaacggttac gatgcgc 780
 ttcacaccaa cgttaacctat cccattacgg tcaatccgc ttttgccttcc acggagaatc 840
 cgacgggttg ttactcgcc acatttatg ttgtgaaag ctggctacag gaaggccaga 900
 60 cgcgaattat ttttgcgttgc gtttaacttgc cgtttcatct gttgtgcac gggcgctgg 960
 tcgggttacgg ccagggacatg cgtttgcgtt ctgaaatttgc cctgagcgc ttttgcgtt 1020
 cccggagaaaa cgcctcgcc gttatgtgtgc tggtttggag tgacggcagt tatctggaa 1080
 atcaggatata tggccggatg agcggcattt tccgtgcacgt ctcgttgcgt cattaaaccga 1140
 ctacacaaat cagcgttttgc catgttgcgc tccgtttaa tgatgatccc agccgcgc 1200
 tactggggcc tgaagttcag atgtgcgcgc agttgcgttgc ctacctacgg gtaacagtt 1260
 ctttatggca gggtaaaacg caggcgcgc gcccgcaccgc gccttcggc ggtgaaat 1320
 tcgatggcc tgggtttat cgcgtcgcc tcacactacg tctgaacgcg gaaaacccga 1380
 65 aactgtggatc cggccaaatc cggaaatctt atcgatcggtt ggttgcactg cacaccgc 1440
 acggcactgtt gatggaaagca gaaacgcgtt gatgcgtt cccgcggatgtt cggattgaa 1500
 atggtctgtt gtcgtgcac ggcgttac ggcgttac cgtcactgc 1560

atcatcctct gcatggtcag gtcatggatg agcagacgat ggtgcaggat atcctgtga 1620
 tgaaggcagaa caactttaac gccgtgcgt gttcgcatta tccgaaccat cgcgtgttgt 1680
 5 acacgctgtg cgaccgctac ggcctgtatg tggggatga agccaatatt gaaaccacg 1740
 gcatggtgc aatgaatcgt ctgaccgatg atccgcgtg gctaccggcg atgagcgaac 1800
 gctgaacgcg aatggtcag cgcgatcgta atcaccgcg agtgcgtatc tggcgctgg 1860
 ggaatgaatc aggccacggc gctaattcact acgcgtgtc tcgctggatc aaatctgtcg 1920
 atccctcccg cccgtgcag tatgaaggcg gcgagccga caccacggcc accgatatta 1980
 tttggccgtat gtaacgcgc gttggatgaag accagccctt cccggctgtg cggaaatgtt 2040
 10 ccatcaaaaa atggcttcg ctaccgtgg agacgcgc gctgatcctt tgcgaatacg 2100
 cccacgcgtat gggtaaacat cttggcggt tcgctaaata ctggcaggcg tttcgtcagt 2160
 atccccgtt acagggcggc ttgcgtctgg actgggtgga tcagtcgtg attaaatatg 2220
 atgaaaacgg caacccgtgg tggcttacg gcggtgatt tggcgatcag cccgaacgatc 2280
 15 gccagttctg tatgaacggc ctggctttg ccgaccgcac gcccgtatcca ggcgtacgg 2340
 aagcaaaaaca ccagcagcag ttttccgtt tccgtttatc cggcaaaacc atcgaagtga 2400
 ccagcgaata cctgttccgt catagcata acgagctct gcactgtatc gtggcgctgg 2460
 atggtaaaggc gctggcaagc ggtgaagtgc ctctggatgt cgctccacaa ggtaaacagt 2520
 20 tgattgaact gctgtacta cggcagccgg agagcgcgg gcaactctgg ctcacagtac 2580
 gctgtgtca accgaacgcg accgcatgtt cagaagccgg gcacatcagc gcctggcagc 2640
 agtggcgtct ggccggaaaac ctcagtgta cgctccccgc cgcgtccac gccatcccg 2700
 25 atctgaccac cagcgaaatg gatttttgc tcgagctggg taataagctg tggcaattt 2760
 accgcccgtc aggctttctt tcacagatgt ggattggcga taaaaaacaat ctgctgacgc 2820
 cgctgcgcga tcagttcacc cgtgcaccgc tggataacga cattggcgta agtgaagcga 2880
 cccgcattga ccctaaccgc tgggtcgaaac gttggaaaggc ggcgggcat taccaggccg 2940
 30 aagcagcgtt gttcgtatgc cccgcgttgc gggaaaacct tatttacgtatc cggaaaacc taccggattt 3060
 atggtagtgg tcaaataggcg gttccgttgc gggccgttgc ggcggcgtat acaccgcata 3120
 cggcgcggat tggctgttgc gttccgttgc gggccgttgc agagcgggta aactggctcg 3180
 35 gatttagggcc gcaagaaaaac tttccgcacc gcttacgcg cgcctgtttt gaccgctggg 3240
 atctgcatt gtcagacatg tatacccccgt tttccgttgc acgtcttccc gagcggaaac ggtctgcgt 3300
 40 gccccgcg cgaattgtat tttccgcacc accagtggcg cggcacttc cagttcaaca 3360
 tcagccgtca cagtcacacg caactgtatc aaaccagcga tcgccatctg ctgcacgcgg 3420
 aagaaggcac atggctgaat atgcacgtt tccatatggg gatgtggc gaccgactt 3480
 ggagccgcgtc agtatcgccg gaaatccagc tgacgcggg tcgcttaccat taccagttgg 3540
 45 tctgggtgtca aaaataataa taaccggca gggggatct ttttgcgtatc ttttgcgtatc 3600
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 3660
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 3720
 50 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 3780
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 3840
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 3900
 55 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 3960
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4020
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4080
 60 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4140
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4200
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4260
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4320
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4380
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4440
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4500
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4560
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4620
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4680
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4740
 65 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4800
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4860
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4920
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 4980
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5040
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5100
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5160
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5220
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5280
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5340
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5400
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5460
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5520
 ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc ttttgcgtatc 5580

14

cagcttggag cgaacgacct acaccgaact gagataccta cagcgtgagc tatgagaag 5640
 cgccacgcctt cccgaaggga gaaaggcgga caggtatccg gtaagcggca gggtcggAAC 5700
 aggagagcgc acgagggagc ttccaggggg aaacgcctgg tatcttata gtcctgtcgg 5760
 5 gtttcgcac ctctgacttg agcgtcgatt tttgtatgc tcgtcaggGG ggcggagcct 5820
 atggaaaaac gccagcaacg cggcctttt acggttcctg gcctttgc ggccttttgc 5880
 tcacatgttc tttcctgcgt tateccctga ttctgtggat aaccgtattt ccgcctttga 5940
 gtgagctgat accgcctcgcc gcagccgaac gaccgagcgc agcgagtca gtagcggagga 6000
 10 agcggaaagag cggccaaatac gcaaaccggc tctccccgcg cgttggccga ttcatatgt 6060
 cagctggcac gacaggtttc cggactggaa agcgggcagt gaggcgaacg caattatgt 6120
 gagtttagctc actcatttagg caccggcggc tttacacttt atgcttccgg ctcgtatgtt 6180
 gtgtggaaatt gtgagcggat aacaatttca cacaggaaac agctatgacc atgattacgc 6240
 15 caagctggcg cg 6252

15 <210> 17
 <211> 25
 <212> DNA
 <213> Artificial Sequence

20 <220>
 <223> Description of Artificial Sequence: primer P64-1

<400> 17
 25 tcagcaacca ggctcccccag cagggc 25

<210> 18
 <211> 27
 <212> DNA
 30 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer 64-4

35 <400> 18
 gacgacagta tcggcctcag gaagatc 27

40 <210> 19
 <211> 840
 <212> DNA
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence:
 oligonucleotide 80d

<400> 19
 50 ggtaccggc tcggatcctc tagtaacggc cgccagtggtc ctggatttcg gttcagcaa 60
 ccaggctccc cagcaggcag aagtatgca agcatgcato tcaattatgc agcaaccagg 120
 tggaaagt cccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag 180
 tcagcaacca tagtcccgcc cctaactccg cccatccgc ccctaactcc gcccagttcc 240
 gcccattctc cggccatgg ctgactaatt tttttatTT atgcagaggg cgaggccggc 300
 tcggcctagg aacagtcgac gacactgcag agacactactt cactaacaac cggtagatTT 360
 55 cgtggaccag atgggtgagg tggagtacgc gcccggggag cccaaaggTT accccagTT 420
 gggcactact cccggaaacc gcttctggat ccataacttc gtatagcata cattatacga 480
 agttataccg ggcacccatg gtgcgagta gcttggact ggggttgcTT ttgcgnnygtc 540
 gtgactggga aaaccctggc gttacccaaac ttaatgcct tgcagcacat ccccccttgc 600
 60 ccagctggcg taatagcga gaggccgcg ccgatgcccc ttcccaacag ttgcgcagct 660
 gaatggcga tggcgctttg cctggctcc ggcaccagaa gcggtgccgg aaagctggct 720
 ggagtgcgat cttcctgagg ccgatactgt cgtaagccg aattctgcag atatccatca 780
 cactggcggc cgctcgagca tgcatactaga gggccaaattt ggcctatagt gagtcgtatt 840

65 <210> 20
 <211> 1842
 <212> DNA

<213> Bacteriophage phi-C31

<220>
 <221> CDS
 5 <222> (1)..(1839)

<400> 20
 atg aca caa ggg gtt gtg acc ggg gtg gac acg tac gcg ggt gct tac 48
 Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr
 10 1 5 10 15

gac cgt cag tcg cgc gag cgc gag aat tcg agc gca gca agc cca gcg 96
 Asp Arg Gln Ser Arg Glu Arg Asn Ser Ser Ala Ala Ser Pro Ala
 15 20 25 30

aca cag cgt agc gcc aac gaa gac aag gcg gcc gac ctt cag cgc gaa 144
 Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gln Arg Glu
 35 40 45

20 gtc gag cgc gac ggg ggc cgg ttc agg ttc gtc ggg cat ttc agc gaa 192
 Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu
 50 55 60

25 gcg ccg ggc acg tcg gcg ttc ggg acg gcg gag cgc ccc gag ttc gaa 240
 Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu
 65 70 75 80

30 cgc atc ctg aac gaa tgc cgc gcc ggg cgg ctc aac atg atc att gtc 288
 Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val
 85 90 95

tat gac gtg tcg cgc ttc tcg cgc ctg aag gtc atg gac gcg att ccg 336
 Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro
 100 105 110

35 att gtc tcg gaa ttg ctc gcc ctg ggc gtg acg att gtt tcc act cag 384
 Ile Val Ser Glu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln
 115 120 125

40 gaa ggc gtc ttc cgg cag gga aac gtc atg gac ctg att cac ctg att 432
 Glu Gly Val Phe Arg Gln Gly Asn Val Met Asp Leu Ile His Leu Ile
 130 135 140

45 atg cgg ctc gac gcg tcg cac aaa gaa tct tcg ctg aag tcg gcg aag 480
 Met Arg Leu Asp Ala Ser His Lys Glu Ser Ser Leu Lys Ser Ala Lys
 145 150 155 160

50 att ctc gac acg aag aac ctt cag cgc gaa ttg ggc ggg tac gtc ggc 528
 Ile Leu Asp Thr Lys Asn Leu Gln Arg Glu Leu Gly Gly Tyr Val Gly
 165 170 175

55 ggg aag gcg cct tac ggc ttc gag ctt gtt tcg gag acg aag gag atc 576
 Gly Lys Ala Pro Tyr Gly Phe Glu Leu Val Ser Glu Thr Lys Glu Ile
 180 185 190

acg cgc aac ggc cga atg gtc aat gtc gtc atc aac aag ctt gcg cac 624
 Thr Arg Asn Gly Arg Met Val Asn Val Val Ile Asn Lys Leu Ala His
 195 200 205

60 tcg acc act ccc ctt acc gga ccc ttc gag ttc gag ccc gac gta atc 672
 Ser Thr Thr Pro Leu Thr Gly Pro Phe Glu Phe Glu Pro Asp Val Ile
 210 215 220

65 cgg tgg tgg tgg cgt gag atc aag acg cac aaa cac ctt ccc ttc aag 720
 Arg Trp Trp Trp Arg Glu Ile Lys Thr His Lys His Leu Pro Phe Lys
 225 230 235 240

ccg	ggc	agt	caa	gcc	gcc	att	cac	ccg	ggc	agc	atc	acg	ggg	ctt	tgt	768	
Pro	Gly	Ser	Gln	Ala	Ala	Ile	His	Pro	Gly	Ser	Ile	Thr	Gly	Leu	Cys		
245								250						255			
5	aag	cgc	atg	gac	gct	gac	gcc	gtg	ccg	acc	ccg	ggc	gag	acg	att	816	
	Lys	Arg	Met	Asp	Ala	Asp	Ala	Val	Pro	Thr	Arg	Gly	Glu	Thr	Ile	Gly	
	260							265						270			
10	aag	aag	acc	gct	tca	agc	gcc	tgg	gac	ccg	gca	acc	gtt	atg	cga	atc	864
	Lys	Lys	Thr	Ala	Ser	Ser	Ala	Trp	Asp	Pro	Ala	Thr	Val	Met	Arg	Ile	
	275							280						285			
15	ctt	ccg	gac	ccg	cgt	att	gcg	ggc	ttc	gcc	gtc	gag	gtg	atc	tac	aag	912
	Leu	Arg	Asp	Pro	Arg	Ile	Ala	Gly	Phe	Ala	Ala	Glu	Val	Ile	Tyr	Lys	
	290						295						300				
20	aag	aag	ccg	gac	ggc	acg	ccg	acc	acg	aag	att	gag	ggt	tac	cgc	att	960
	Lys	Lys	Pro	Asp	Gly	Thr	Pro	Thr	Thr	Lys	Ile	Glu	Gly	Tyr	Arg	Ile	
	305						310				315			320			
25	cag	cgc	gac	ccg	atc	acg	ctc	ccg	ccg	gtc	gag	ctt	gat	tgc	gga	ccg	1008
	Gln	Arg	Asp	Pro	Ile	Thr	Leu	Arg	Pro	Val	Glu	Leu	Asp	Cys	Gly	Pro	
	325						330				335			335			
30	atc	atc	gag	ccc	gct	gag	tgg	tat	gag	ctt	cag	gcg	tgg	ttg	gac	ggc	1056
	Ile	Ile	Glu	Pro	Ala	Glu	Trp	Tyr	Glu	Leu	Gln	Ala	Trp	Leu	Asp	Gly	
	340						345							350			
35	agg	ggg	cgc	ggc	aag	ggg	ctt	tcc	cg	gg	caa	gcc	att	ctg	tcc	ggc	1104
	Arg	Gly	Arg	Gly	Lys	Gly	Leu	Ser	Arg	Gly	Gln	Ala	Ile	Leu	Ser	Ala	
	355						360							365			
40	atg	gac	aag	ctg	tac	tgc	gag	tgt	ggc	gcc	gtc	atg	act	tgc	aag	cgc	1152
	Met	Asp	Lys	Leu	Tyr	Cys	Glu	Cys	Gly	Ala	Val	Met	Thr	Ser	Lys	Arg	
	370						375							380			
45	ggg	gaa	gaa	tcg	atc	aag	gac	tct	tac	ccg	tgc	cgt	cgc	ccg	aag	gtg	1200
	Gly	Glu	Glu	Ser	Ile	Lys	Asp	Ser	Tyr	Arg	Cys	Arg	Arg	Arg	Lys	Val	
	385						390				395			400			
50	gtc	gac	ccg	tcc	gca	cct	ggg	cag	cac	gaa	ggc	acg	tgc	aac	gtc	agc	1248
	Val	Asp	Pro	Ser	Ala	Pro	Gly	Gln	His	Glu	Gly	Thr	Cys	Asn	Val	Ser	
	405						410							415			
55	atg	gcg	gca	ctc	gac	aag	ttc	gtt	gcg	gaa	cgc	atc	ttc	aac	aag	atc	1296
	Met	Ala	Ala	Leu	Asp	Lys	Phe	Val	Ala	Glu	Arg	Ile	Phe	Asn	Lys	Ile	
	420						425							430			
60	agg	cac	gcc	gaa	ggc	gac	gaa	gag	acg	ttg	gcg	ctt	ctg	tgg	gaa	gcc	1344
	Arg	His	Ala	Glu	Gly	Asp	Glu	Glu	Thr	Leu	Ala	Leu	Leu	Trp	Glu	Ala	
	435						440							445			
65	gcc	cga	cgc	tcc	ggc	aag	ctc	act	gag	gcg	cct	gag	aag	agc	ggc	gaa	1392
	Ala	Arg	Arg	Phe	Gly	Lys	Leu	Thr	Glu	Ala	Pro	Glu	Lys	Ser	Gly	Glu	
	450						455							460			
70	cgg	gcg	aac	ctt	gtt	gcg	gag	ccg	gac	gcc	ctg	aac	gcc	ctt	gaa		1440
	Arg	Ala	Asn	Leu	Val	Ala	Glu	Arg	Ala	Asp	Ala	Leu	Asn	Ala	Leu	Glu	
	465						470				475			480			
75	gag	ctg	tac	gaa	gac	ccg	gca	ggc	gog	tac	gac	gga	ccc	gtt	ggc		1488
	Glu	Leu	Tyr	Glu	Asp	Arg	Ala	Ala	Gly	Ala	Tyr	Asp	Gly	Pro	Val	Gly	
	485						490				495			495			
80	agg	aag	cac	tcc	ccg	aag	caa	cag	gca	gca	ctg	acg	ctc	ccg	cag	caa	1536
	Arg	Lys	His	Phe	Arg	Lys	Gln	Gln	Ala	Ala	Leu	Thr	Leu	Arg	Gln	Gln	
	500						505							510			

ggg gcg gaa gag cgg ctt gcc gaa ctt gaa gcc gcc gaa gcc ccg aag 1584
 Gly Ala Glu Glu Arg Leu Ala Glu Leu Glu Ala Ala Glu Ala Pro Lys
 515 520 525

5 ctt ccc ctt gac caa tgg ttc ccc gaa gac gcc gac gct gac ccg acc 1632
 Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr
 530 535 540

10 ggc cct aag tcg tgg tgg ggg cgc gcg tca gta gac gac aag cgc gtg 1680
 Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val
 545 550 555 560

15 ttc gtc ggg ctc ttc gta gac aag atc gtt gtc acg aag tcg act acg 1728
 Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr
 565 570 575

20 ggc agg ggg cag gga acg ccc atc gag aag cgc gct tcg atc acg tgg 1776
 Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp
 580 585 590

25 gcg aag ccg ccg acc gac gac gac gaa gac gac gcc cag gac ggc acg 1824
 Ala Lys Pro Pro Thr Asp Asp Asp Glu Asp Asp Ala Gln Asp Gly Thr
 595 600 605

30 gaa gac gta gcg gcg tag 1842
 Glu Asp Val Ala Ala
 610

35 <210> 21
 <211> 613
 <212> PRT
 <213> Bacteriophage phi-C31

40 <400> 21
 Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr
 1 5 10 15

45 Asp Arg Gln Ser Arg Glu Arg Glu Asn Ser Ser Ala Ala Ser Pro Ala
 20 25 30

Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gln Arg Glu
 35 40 45

50 Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu
 50 55 60

55 Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu
 65 70 75 80

Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val
 85 90 95

60 Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro
 100 105 110

Ile Val Ser Glu Leu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln
 115 120 125

65 Glu Gly Val Phe Arg Gln Gly Asn Val Met Asp Leu Ile His Leu Ile
 130 135 140

Met Arg Leu Asp Ala Ser His Lys Glu Ser Ser Leu Lys Ser Ala Lys
 145 150 155 160

Ile Leu Asp Thr Lys Asn Leu Gln Arg Glu Leu Gly Gly Tyr Val Gly

		18		
	165	170	175	
	Gly Lys Ala Pro Tyr Gly Phe Glu Leu Val Ser Glu Thr Lys Glu Ile			
	180	185	190	
5	Thr Arg Asn Gly Arg Met Val Asn Val Val Ile Asn Lys Leu Ala His			
	195	200	205	
	Ser Thr Thr Pro Leu Thr Gly Pro Phe Glu Phe Glu Pro Asp Val Ile			
10	210	215	220	
	Arg Trp Trp Trp Arg Glu Ile Lys Thr His Lys His Leu Pro Phe Lys			
	225	230	235	240
15	Pro Gly Ser Gln Ala Ala Ile His Pro Gly Ser Ile Thr Gly Leu Cys			
	245	250	255	
	Lys Arg Met Asp Ala Asp Ala Val Pro Thr Arg Gly Glu Thr Ile Gly			
20	260	265	270	
	Lys Lys Thr Ala Ser Ser Ala Trp Asp Pro Ala Thr Val Met Arg Ile			
	275	280	285	
25	Leu Arg Asp Pro Arg Ile Ala Gly Phe Ala Ala Glu Val Ile Tyr Lys			
	290	295	300	
	Lys Lys Pro Asp Gly Thr Pro Thr Thr Lys Ile Glu Gly Tyr Arg Ile			
	305	310	315	320
30	Gln Arg Asp Pro Ile Thr Leu Arg Pro Val Glu Leu Asp Cys Gly Pro			
	325	330	335	
	Ile Ile Glu Pro Ala Glu Trp Tyr Glu Leu Gln Ala Trp Leu Asp Gly			
35	340	345	350	
	Arg Gly Arg Gly Lys Gly Leu Ser Arg Gly Gln Ala Ile Leu Ser Ala			
	355	360	365	
40	Met Asp Lys Leu Tyr Cys Glu Cys Gly Ala Val Met Thr Ser Lys Arg			
	370	375	380	
	Gly Glu Glu Ser Ile Lys Asp Ser Tyr Arg Cys Arg Arg Arg Lys Val			
	385	390	395	400
45	Val Asp Pro Ser Ala Pro Gly Gln His Glu Gly Thr Cys Asn Val Ser			
	405	410	415	
	Met Ala Ala Leu Asp Lys Phe Val Ala Glu Arg Ile Phe Asn Lys Ile			
50	420	425	430	
	Arg His Ala Glu Gly Asp Glu Glu Thr Leu Ala Leu Leu Trp Glu Ala			
	435	440	445	
55	Ala Arg Arg Phe Gly Lys Leu Thr Glu Ala Pro Glu Lys Ser Gly Glu			
	450	455	460	
	Arg Ala Asn Leu Val Ala Glu Arg Ala Asp Ala Leu Asn Ala Leu Glu			
	465	470	475	480
60	Glu Leu Tyr Glu Asp Arg Ala Ala Gly Ala Tyr Asp Gly Pro Val Gly			
	485	490	495	
	Arg Lys His Phe Arg Lys Gln Gln Ala Ala Leu Thr Leu Arg Gln Gln			
65	500	505	510	
	Gly Ala Glu Glu Arg Leu Ala Glu Leu Glu Ala Ala Glu Ala Pro Lys			
	515	520	525	

Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr
 530 535 540

5 Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val
 545 550 555 560

Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr
 565 570 575

10 Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp
 580 585 590

Ala Lys Pro Pro Thr Asp Asp Asp Glu Asp Asp Ala Gln Asp Gly Thr
 15 595 600 605

Glu Asp Val Ala Ala
 610

20

<210> 22
 <211> 1863
 <212> DNA
 25 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: DNA sequence
 coding for fusion protein C31-Int(CNLS)

30 <220>
 <221> CDS
 <222> (1)..(1860)

35 <400> 22
 atg aca caa ggg gtt gtg acc ggg gtg gac acg tac gcg ggt gct tac 48
 Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr
 1 5 10 15

40 gac cgt cag tcg cgc gag cgc gag aat tcg agc gca gca agc cca gcg 96
 Asp Arg Gln Ser Arg Glu Arg Glu Asn Ser Ser Ala Ala Ser Pro Ala
 20 25 30

45 aca cag cgt agc gcc aac gaa gac aag gcg gcc gac ctt cag cgc gaa 144
 Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gin Arg Glu
 35 40 45

50 gtc gag cgc gac ggg ggc cgg ttc agg ttc gtc ggg cat ttc agc gaa 192
 Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu
 50 55 60

55 gcg ccc ggc acg tcg gcg ttc ggg acg gcg gag cgc ccc gag ttc gaa 240
 Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu
 65 70 75 80

60 cgc atc ctg aac gaa tgc cgc gcc ggg cgg ctc aac atg atc att gtc 288
 Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val
 85 90 95

65 tat gac gtg tcg cgc ttc tcg cgc ctg aag gtc atg gac gcg att ccg 336
 Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro
 100 105 110

att gtc tcg gaa ttg ctc gcc ctg ggc gtg acg att gtt tcc act cag 384
 Ile Val Ser Glu Leu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln
 115 120 125

20

gaa	ggc	gtc	tcc	cg	cag	gga	aac	gtc	atg	gac	ctg	att	ca	ctg	att	432	
Glu	Gly	Val	Phe	Arg	Gln	Gly	Asn	Val	Met	Asp	Leu	Ile	His	Leu	Ile		
130																	
									135								
5	atg	cg	ctc	gac	gc	tc	c	aa	gaa	tct	tc	ctg	aag	tc	gc	aag	480
Met	Arg	Leu	Asp	Ala	Ser	His	Lys	Glu	Ser	Ser	Leu	Lys	Ser	Ala	Lys		
145										150							
10	att	ctc	gac	ac	aa	gaa	aa	ttg	ggc	gg	ta	gc	gg			528	
Ile	Leu	Asp	Thr	Lys	Asn	Leu	Gln	Glu	Leu	Gly	Gly	Tyr	Val	Gly			
									165			170			175		
15	ggg	aag	gc	cct	ta	ggc	ttc	ga	tt	tc	ga	ac	aa	ga	at	576	
Gly	Lys	Ala	Pro	Tyr	Gly	Phe	Glu	Leu	Val	Ser	Glu	Thr	Lys	Glu	Ile		
									180			185			190		
20	ac	cg	aa	gg	cg	at	gt	aa	gt	gt	at	aa	ct	gg	ca	624	
Thr	Arg	Asn	Gly	Arg	Met	Val	Asn	Val	Val	Ile	Asn	Lys	Leu	Ala	His		
									195			200			205		
25	tc	ac	ac	cc	ct	ac	gg	cc	ga	cc	ga	cc	gg	ca	at	672	
Ser	Thr	Thr	Pro	Leu	Thr	Gly	Pro	Phe	Glu	Phe	Glu	Pro	Asp	Val	Ile		
									210			215			220		
30	cg	tg	tg	tg	cg	tg	aa	ac	ca	aa	ca	ct	cc	tt	aa	720	
Arg	Trp	Trp	Trp	Arg	Glu	Ile	Lys	Thr	His	Lys	His	Leu	Pro	Phe	Lys		
									225			230			240		
35	cc	gg	ag	ca	gg	cc	at	ca	cc	gg	ac	at	cc	gg	ctt	768	
Pro	Gly	Ser	Gln	Ala	Ala	Ile	His	Pro	Gly	Ser	Ile	Thr	Gly	Leu	Cys		
									245			250			255		
40	aa	cg	at	g	ct	g	cc	gt	cc	ac	cg	gg	ac	at	gg	816	
Lys	Arg	Met	Asp	Ala	Asp	Ala	Val	Pro	Thr	Arg	Gly	Glu	Thr	Ile	Gly		
									260			265			270		
45	aa	aa	ac	g	ct	g	cc	tt	cc	ac	tt	at	cg	at	864		
Lys	Lys	Pro	Asp	Gly	Thr	Pro	Thr	Lys	Ile	Glu	Gly	Tyr	Arg	Ile			
									305			310			320		
50	ca	cg	ga	cc	at	ac	ct	cg	gg	ct	gt	tg	gg	cc	1008		
Gln	Arg	Asp	Pro	Ile	Thr	Leu	Arg	Pro	Val	Glu	Leu	Asp	Cys	Gly	Pro		
									325			330			335		
55	at	at	ga	cc	gt	ga	tg	ta	ga	ct	ca	gc	tg	tt	ga	956	
Ile	Ile	Glu	Pro	Ala	Glu	Trp	Tyr	Glu	Leu	Gln	Ala	Trp	Leu	Asp	Gly		
									340			345			350		
60	ag	gg	cg	gg	aa	gg	ct	cc	cg	gg	ca	aa	cc	tt	gg	1104	
Arg	Gly	Arg	Gly	Lys	Gly	Leu	Ser	Arg	Gly	Gln	Ala	Ile	Leu	Ser	Ala		
									355			360			365		
65	at	g	aa	ct	ta	tc	g	tg	tg	gg	cc	at	tc	aa	cg	1152	
Met	Asp	Lys	Leu	Tyr	Cys	Glu	Cys	Gly	Ala	Val	Met	Thr	Ser	Lys	Arg		
									370			375			380		
70	gg	ga	aa	tc	at	aa	g	tc	tc	cg	tg	cg	cc	aa	gt	1200	
Gly	Glu	Glu	Ser	Ile	Lys	Asp	Ser	Tyr	Arg	Cys	Arg	Arg	Arg	Lys	Val		
									385			390			395		

5	gtc gac ccg tcc gca cct ggg cag cac gaa ggc acg tgc aac gtc agc Val Asp Pro Ser Ala Pro Gly Gln His Glu Gly Thr Cys Asn Val Ser 405 410 415	1248
10	atg gcg gca ctc gac aag ttc gtt gcg gaa cgc atc ttc aac aag atc Met Ala Ala Leu Asp Lys Phe Val Ala Glu Arg Ile Phe Asn Lys Ile 420 425 430	1296
15	agg cac gcc gaa ggc gac gaa gag acg ttg gcg ctt ctg tgg gaa gcc Arg His Ala Glu Gly Asp Glu Glu Thr Leu Ala Leu Leu Trp Glu Ala 435 440 445	1344
20	gcc cga cgc ttc ggc aag ctc act gag gcg cct gag aag acg ggc gaa Ala Arg Arg Phe Gly Lys Leu Thr Glu Ala Pro Glu Lys Ser Gly Glu 450 455 460	1392
25	cgg gcg aac ctt gtt gcg gag cgc gcc gac gcc ctg aac gcc ctt gaa Arg Ala Asn Leu Val Ala Glu Arg Ala Asp Ala Leu Asn Ala Leu Glu 465 470 475 480	1440
30	gag ctg tac gaa gac cgc gca ggc gcg tac gac gga ccc gtt ggc Glu Leu Tyr Glu Asp Arg Ala Ala Gly Ala Tyr Asp Gly Pro Val Gly 485 490 495	1488
35	agg aag cac ttc cgg aag caa cag gca gcg ctg acg ctc cgg cag caa Arg Lys His Phe Arg Lys Gln Gln Ala Ala Leu Thr Leu Arg Gln Gln 500 505 510	1536
40	ggg gcg gaa gag cgg ctt gcc gaa ctt gaa gcc gcc gaa gcc ccc aag Gly Ala Glu Glu Arg Leu Ala Glu Leu Glu Ala Ala Pro Lys 515 520 525	1584
45	ctt ccc ctt gac caa tgg ttc ccc gaa gac gcc gac gct gac ccc acc Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr 530 535 540	1632
50	ggc cct aag tcg tgg tgg ggg cgc gcg tca gta gac gac aag cgc gtg Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val 545 550 555 560	1680
55	tcc gtc ggg ctc ttc gta gac aag atc gtt gtc acg aag tcg act acg Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr 565 570 575	1728
60	ggc agg ggg cag gga acg ccc atc gag aag cgc gct tcg atc acg tgg Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp 580 585 590	1776
65	gaa gac gta gcg gcg cct aag aag aag agg aag gtt tag Glu Asp Val Ala Ala Pro Lys Lys Lys Arg Lys Val 610 615 620	1863
70	<210> 23 <211> 620 <212> PRT <213> Artificial Sequence <223> Description of Artificial Sequence: DNA sequence coding for fusion protein C31-Int(CNLS)	
75	<400> 23 Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr	

		22		
1	5	10		15
Asp Arg Gln Ser Arg Glu Arg Glu Asn Ser Ser Ala Ala Ser Pro Ala				
	20	25		30
5	Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gln Arg Glu			
	35	40		45
10	Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu			
	50	55		60
Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu				
	65	70		75
15	Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val			
	85	90		95
20	Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro			
	100	105		110
Ile Val Ser Glu Leu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln				
	115	120		125
25	Glu Gly Val Phe Arg Gln Gly Asn Val Met Asp Leu Ile His Leu Ile			
	130	135		140
Met Arg Leu Asp Ala Ser His Lys Glu Ser Ser Leu Lys Ser Ala Lys				
	145	150		155
30	Ile Leu Asp Thr Lys Asn Leu Gln Arg Glu Leu Gly Gly Tyr Val Gly			
	165	170		175
Gly Lys Ala Pro Tyr Gly Phe Glu Leu Val Ser Glu Thr Lys Glu Ile				
	180	185		190
35	Thr Arg Asn Gly Arg Met Val Asn Val Val Ile Asn Lys Leu Ala His			
	195	200		205
40	Ser Thr Thr Pro Leu Thr Gly Pro Phe Glu Phe Glu Pro Asp Val Ile			
	210	215		220
Arg Trp Trp Trp Arg Glu Ile Lys Thr His Lys His Leu Pro Phe Lys				
	225	230		235
240				
45	Pro Gly Ser Gln Ala Ala Ile His Pro Gly Ser Ile Thr Gly Leu Cys			
	245	250		255
Lys Arg Met Asp Ala Asp Ala Val Pro Thr Arg Gly Glu Thr Ile Gly				
	260	265		270
50	Lys Lys Thr Ala Ser Ser Ala Trp Asp Pro Ala Thr Val Met Arg Ile			
	275	280		285
Leu Arg Asp Pro Arg Ile Ala Gly Phe Ala Ala Glu Val Ile Tyr Lys				
	290	295		300
55	Lys Lys Pro Asp Gly Thr Pro Thr Thr Lys Ile Glu Gly Tyr Arg Ile			
	305	310		315
320				
60	Gln Arg Asp Pro Ile Thr Leu Arg Pro Val Glu Leu Asp Cys Gly Pro			
	325	330		335
Ile Ile Glu Pro Ala Glu Trp Tyr Glu Leu Gln Ala Trp Leu Asp Gly				
	340	345		350
65	Arg Gly Arg Gly Lys Gly Leu Ser Arg Gly Gln Ala Ile Leu Ser Ala			
	355	360		365

Met Asp Lys Leu Tyr Cys Glu Cys Gly Ala Val Met Thr Ser Lys Arg
 370 375 380

5 Gly Glu Glu Ser Ile Lys Asp Ser Tyr Arg Cys Arg Arg Arg Lys Val
 385 390 395 400

Val Asp Pro Ser Ala Pro Gly Gln His Glu Gly Thr Cys Asn Val Ser
 405 410 415

10 Met Ala Ala Leu Asp Lys Phe Val Ala Glu Arg Ile Phe Asn Lys Ile
 420 425 430

15 Arg His Ala Glu Gly Asp Glu Glu Thr Leu Ala Leu Leu Trp Glu Ala
 435 440 445

Ala Arg Arg Phe Gly Lys Leu Thr Glu Ala Pro Glu Lys Ser Gly Glu
 450 455 460

20 Arg Ala Asn Leu Val Ala Glu Arg Ala Asp Ala Leu Asn Ala Leu Glu
 465 470 475 480

Glu Leu Tyr Glu Asp Arg Ala Ala Gly Ala Tyr Asp Gly Pro Val Gly
 485 490 495

25 Arg Lys His Phe Arg Lys Gln Gln Ala Ala Leu Thr Leu Arg Gln Gln
 500 505 510

30 Gly Ala Glu Glu Arg Leu Ala Glu Leu Glu Ala Ala Glu Ala Pro Lys
 515 520 525

Leu Pro Leu Asp Gln Trp Phe Pro Glu Asp Ala Asp Ala Asp Pro Thr
 530 535 540

35 Gly Pro Lys Ser Trp Trp Gly Arg Ala Ser Val Asp Asp Lys Arg Val
 545 550 555 560

Phe Val Gly Leu Phe Val Asp Lys Ile Val Val Thr Lys Ser Thr Thr
 565 570 575

40 Gly Arg Gly Gln Gly Thr Pro Ile Glu Lys Arg Ala Ser Ile Thr Trp
 580 585 590

45 Ala Lys Pro Pro Thr Asp Asp Glu Asp Asp Ala Gln Asp Gly Thr
 595 600 605

Glu Asp Val Ala Ala Pro Lys Lys Lys Arg Lys Val
 610 615 620

50

<210> 24
 <211> 43
 <212> PRT
 55 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: NLS

60 <400> 24
 Met Lys Lys Lys Lys Lys Lys Lys Lys Lys Cys Arg Leu Lys
 1 5 10 15

65 Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu Lys
 20 25 30

Lys Lys Lys Arg Arg Lys Thr Lys Arg

35

40

5 <210> 25
<211> 10
<212> PRT
<213> Artificial Sequence

10 <220>
10 <223> Description of Artificial Sequence: NLS

15 <400> 25
Ile Lys Tyr Phe Lys Lys Phe Pro Lys Asp
1 5 10

20 <210> 26
<211> 14
<212> PRT
20 <213> Artificial Sequence

25 <220>
25 <223> Description of Artificial Sequence: NLS

30 <400> 26
Met Thr Gly Ser Lys Thr Arg Lys His Arg Gly Ser Gly Ala
1 5 10

35 <210> 27
<211> 14
<212> PRT
<213> Artificial Sequence

40 <220>
40 <223> Description of Artificial Sequence: NLS

45 <400> 27
Met Thr Gly Ser Lys His Arg Lys His Pro Gly Ser Gly Ala
1 5 10

50 <210> 28
<211> 7
<212> PRT
50 <213> Artificial Sequence

55 <220>
55 <223> Description of Artificial Sequence: NLS

60 <400> 28
Gly Lys Lys Arg Ser Lys Ala
1 5

65 <210> 29
<211> 14
<212> PRT
65 <213> Artificial Sequence

65 <220>
65 <223> Description of Artificial Sequence: NLS

65 <400> 29
Pro Lys Lys Ala Arg Glu Asp Val Ser Arg Lys Arg Pro Arg
1 5 10

5 <210> 30
 <211> 11
 <212> PRT
 5 <213> Artificial Sequence

10 <220>
 <223> Description of Artificial Sequence: NLS

10 <400> 30
 Ala Pro Lys Arg Lys Ser Gly Val Ser Lys Cys
 1 5 10

15 <210> 31
 <211> 12
 <212> PRT
 5 <213> Artificial Sequence

20 <220>
 <223> Description of Artificial Sequence: NLS

20 <400> 31
 Glu Glu Asp Gly Pro Gln Lys Lys Lys Arg Arg Leu
25 1 5 10

30 <210> 32
 <211> 8
 <212> PRT
 5 <213> Artificial Sequence

35 <220>
 <223> Description of Artificial Sequence: NLS

35 <400> 32
 Ala Pro Thr Lys Arg Lys Gly Ser
 1 5

40 <210> 33
 <211> 7
 <212> PRT
 5 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: NLS

50 <400> 33
 Pro Asn Lys Lys Lys Arg Lys
 1 5

55 <210> 34
 <211> 5
 <212> PRT
 5 <213> Artificial Sequence

60 <220>
 <223> Description of Artificial Sequence: NLS

60 <400> 34
 Lys Arg Pro Arg Pro
 1 5
65 <210> 35

<211> 11
<212> PRT
<213> Artificial Sequence

5 <220>
<223> Description of Artificial Sequence: NLS

10 <400> 35
Cys Gly Gly Leu Ser Ser Lys Arg Pro Arg Pro
1 5 10

15 <210> 36
<211> 19
<212> PRT
<213> Artificial Sequence

20 <220>
<223> Description of Artificial Sequence: NLS

<400> 36
Pro Pro Lys Lys Arg Met Arg Arg Arg Ile Glu Pro Lys Lys Lys
1 5 10 15

25 Lys Arg Pro

30 <210> 37
<211> 11
<212> PRT
<213> Artificial Sequence

35 <220>
<223> Description of Artificial Sequence: NLS

<400> 37
Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys
1 5 10

40 <210> 38
<211> 9
<212> PRT
<213> Artificial Sequence

45 <220>
<223> Description of Artificial Sequence: NLS

50 <400> 38
Pro Lys Gln Lys Arg Lys Met Ala Arg
1 5

55 <210> 39
<211> 9
<212> PRT
<213> Artificial Sequence

60 <220>
<223> Description of Artificial Sequence: NLS

65 <400> 39
Ser Val Thr Lys Lys Arg Lys Leu Glu
1 5

5 <210> 40
 <211> 11
 <212> PRT
 <213> Artificial Sequence

10 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 40
 Cys Gly Gly Ala Ala Lys Arg Val Lys Leu Asp
 1 5 10

15 <210> 41
 <211> 9
 <212> PRT
 <213> Artificial Sequence

20 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 41
 Pro Ala Ala Lys Arg Val Lys Leu Asp
 1 5

25

30 <210> 42
 <211> 11
 <212> PRT
 <213> Artificial Sequence

35 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 42
 Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
 1 5 10

40 <210> 43
 <211> 8
 <212> PRT
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 43
 Pro Gln Ser Arg Lys Lys Leu Arg
 1 5

50

55 <210> 44
 <211> 8
 <212> PRT
 <213> Artificial Sequence

60 <220>
 <223> Description of Artificial Sequence: NLS
 <400> 44
 Pro Leu Leu Lys Lys Ile Lys Gln
 1 5

65 <210> 45
 <211> 7

<212> PRT
<213> Artificial Sequence

5 <220>
<223> Description of Artificial Sequence: NLS

<400> 45
Pro Gln Pro Lys Lys Lys Pro
1 5

10

<210> 46
<211> 9
<212> PRT
15 <213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: NLS

20 <400> 46
Ser Lys Arg Val Ala Lys Arg Lys Leu
1 5

25 <210> 47
<211> 9
<212> PRT
<213> Artificial Sequence

30 <220>
<223> Description of Artificial Sequence: NLS

<400> 47
Ala Ser Lys Ser Arg Lys Arg Lys Leu
1 5

35

<210> 48
<211> 16
40 <212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: NLS

45 <400> 48
Gly Gly Leu Cys Ser Ala Arg Leu His Arg His Ala Leu Leu Ala Thr
1 5 10 15

50

<210> 49
<211> 8
<212> PRT
55 <213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: NLS

<400> 49
60 Arg Lys Thr Lys Lys Lys Ile Lys
1 5

65 <210> 50
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: NLS

5 <400> 50
Arg Lys Leu Lys Lys Leu Gly Asn
1 5

10 <210> 51
<211> 8
<212> PRT
<213> Artificial Sequence

15 <220>
<223> Description of Artificial Sequence: NLS

<400> 51
Arg Lys Asp Arg Arg Gly Gly Arg
1 5

25 <210> 52
<211> 18
<212> PRT
<213> Artificial Sequence

30 <220>
<223> Description of Artificial Sequence: NLS

<400> 52
Asp Thr Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg Leu Leu Arg Leu
1 5 10 15

35 Asp Glu

40 <210> 53
<211> 7
<212> PRT
<213> Artificial Sequence

45 <220>
<223> Description of Artificial Sequence: NLS

<400> 53
Pro Lys Lys Lys Arg Lys Val
1 5

50 <210> 54
<211> 1410
<212> DNA
<213> Bacteriophage R4

55 <220>
<221> CDS
<222> (1)..(1407)

60 <400> 54
atg aat cga ggg ggg ccc act gta cgg gcc gac atc tac gtc cga atc 48
Met Asn Arg Gly Gly Pro Thr Val Arg Ala Asp Ile Tyr Val Arg Ile
1 5 10 15

65 agc ctg gac cgc aca ggg gaa gag ctc ggg gtc gag cgc cag gag gag 96
Ser Leu Asp Arg Thr Gly Glu Glu Leu Gly Val Glu Arg Gln Glu Glu

	20	25	30	
			30	
5	tcg tgt cgc gag ctc tgc aag agc ctc ggc atg gag gtg ggg cag gtg Ser Cys Arg Glu Leu Cys Lys Ser Leu Gly Met Glu Val Gly Gln Val 35 40 45			144
10	tgg gtc gac aac gac ctg agc gcc acc aag aac gtc gtc cgc cct Trp Val Asp Asn Asp Leu Ser Ala Thr Lys Lys Asn Val Val Arg Pro 50 55 60			192
15	gac ttc gag gcg atg atc gcg agc aac ccg cag gcg atc gtc tgc tgg Asp Phe Glu Ala Met Ile Ala Ser Asn Pro Gln Ala Ile Val Cys Trp 65 70 75 80			240
20	cac acc gac cgg ctc atc cgc gtc acg ccg gac ctg gag cgg gtg atc His Thr Asp Arg Leu Ile Arg Val Thr Arg Asp Leu Glu Arg Val Ile 85 90 95			288
25	gac ctc gga gtc aac gtc cac gcc gtg atg gcc gga cac ctg gac ctg Asp Leu Gly Val Asn Val His Ala Val Met Ala Gly His Leu Asp Leu 100 105 110			336
30	tcc acc ccg gcc ggc cga gcc gtc gcc cgc acg gtg acg gcc tgg gcc Ser Thr Pro Ala Gly Arg Ala Val Ala Arg Thr Val Thr Ala Trp Ala 115 120 125			384
35	acg tac gag ggc gag cag aag gct gag cgc cag aag ctc gcc aac atc Thr Tyr Glu Gly Glu Gln Lys Ala Glu Arg Gln Lys Leu Ala Asn Ile 130 135 140			432
40	cag aac gcc cgc gcc ggc aag ccg tac acc ccc ggc atc cgc ccc ttc Gln Asn Ala Arg Ala Gly Lys Pro Tyr Thr Pro Gly Ile Arg Pro Phe 145 150 155 160			480
45	ggg tac ggc gac gac cac atg acc atc gtg acg gcc gag gcg gac gcc Gly Tyr Gly Asp Asp His Met Thr Ile Val Thr Ala Glu Ala Asp Ala 165 170 175			528
50	atc cgc gac ggc gcg aag atg atc ctc gac ggc tgg tcc ctg tcg gcc Ile Arg Asp Gly Ala Lys Met Ile Leu Asp Gly Trp Ser Leu Ser Ala 180 185 190			576
55	gtg gct cgc tac tgg gag gag ctc aag ctc cag tcg ccc cgg agt atg Val Ala Arg Tyr Trp Glu Leu Lys Leu Gln Ser Pro Arg Ser Met 195 200 205			624
60	gcc gca ggc ggc aag ggc tgg tct ctg cgg ggc gta aag aag gtg ctg Ala Ala Gly Gly Lys Gly Trp Ser Leu Arg Gly Val Lys Lys Val Leu 210 215 220			672
65	acc tcc ccg cgc tac gtc ggg cgg tcc agc tac ctc ggg gag gtc gtg Thr Ser Pro Arg Tyr Val Gly Arg Ser Ser Tyr Leu Gly Glu Val Val 225 230 235 240			720
70	ggc gat gct cag tgg ccg ccc atc ctc gac ccg gac gtc tac tac ggg Gly Asp Ala Gln Trp Pro Pro Ile Leu Asp Pro Asp Val Tyr Tyr Gly 245 250 255			768
75	gtc gtg gcc atc ctg aac aac ccc gac cgc ttc agc ggg ggc cct cgg Val Val Ala Ile Leu Asn Asn Pro Asp Arg Phe Ser Gly Gly Pro Arg 260 265 270			816
80	acc ggc cgc acc ccc ggc acg ctg ctc gca ggc atc gcc ttg tgc ggt Thr Gly Arg Thr Pro Gly Thr Leu Leu Ala Gly Ile Ala Leu Cys Gly 275 280 285			864
85	gag tgc ggc aag acg gtc agt gga cgc ggc tac cga ggt gtc ctg gtc			912

Glu Cys Gly Lys Thr Val Ser Gly Arg Gly Tyr Arg Gly Val Leu Val
 290 295 300

5 tac gga tgt aag gac acg cac act cgg acg cct cgg agc atc gct gac 960
 Tyr Gly Cys Lys Asp Thr His Thr Arg Thr Pro Arg Ser Ile Ala Asp
 305 310 315 320

10 ggc cgc gcg agc agc tcg acc ctc gcc cgg ctc atg ttc ccc gac ttc 1008
 Gly Arg Ala Ser Ser Thr Leu Ala Arg Leu Met Phe Pro Asp Phe
 325 330 335

15 ctg ccc ggc ctc ctg gcc tct ggg cag gcc gag gac ggc cag tcg gca 1056
 Leu Pro Gly Leu Leu Ala Ser Gly Gln Ala Glu Asp Gly Gln Ser Ala
 340 345 350

20 gca tcc aag cac tcg gag gcc cag acg ctg cgc gag cgc ctt gac ggg 1104
 Ala Ser Lys His Ser Glu Ala Gln Thr Leu Arg Glu Arg Leu Asp Gly
 355 360 365

25 ctg gct acg gcc tac gcg gag ggt gcg atc acg ctg tct cag atg acg 1152
 Leu Ala Thr Ala Tyr Ala Glu Gly Ala Ile Ser Leu Ser Gln Met Thr
 370 375 380

30 gcc ggc tcg gaa gca ctg cgg aag aag ctg gag gtg atc gaa gcc gac 1200
 Ala Gly Ser Glu Ala Leu Arg Lys Lys Leu Glu Val Ile Glu Ala Asp
 385 390 395 400

35 ctc gtg ggc tcg gca ggc atc ccc ttc gat cca gtg gcc gga gtg 1248
 Leu Val Gly Ser Ala Gly Ile Pro Pro Phe Asp Pro Val Ala Gly Val
 405 410 415

40 gct ggc ctg atc tcc ggc tgg ccc acc acg cct ctc ccc acg cgt cga 1296
 Ala Gly Leu Ile Ser Gly Trp Pro Thr Thr Pro Leu Pro Thr Arg Arg
 420 425 430

45 gca tgg gtg gac ttc tgc ctg gtg gtc acg ctg aac acc cag aag ggg 1344
 Ala Trp Val Asp Phe Cys Leu Val Val Thr Leu Asn Thr Gln Lys Gly
 435 440 445

50 cgc cat gcg tcg agc atg acc gtg gac gac cac gtc acc atc gag tgg 1392
 Arg His Ala Ser Ser Met Thr Val Asp Asp His Val Thr Ile Glu Trp
 450 455 460

55 cga gac gtg gcc gag tag 1410
 Arg Asp Val Ala Glu
 465

60 <210> 55
 <211> 469
 <212> PRT
 <213> Bacteriophage R4

65 <400> 55
 Met Asn Arg Gly Gly Pro Thr Val Arg Ala Asp Ile Tyr Val Arg Ile
 1 5 10 15

Ser Leu Asp Arg Thr Gly Glu Leu Gly Val Glu Arg Gln Glu Glu
 20 25 30

Ser Cys Arg Glu Leu Cys Lys Ser Leu Gly Met Glu Val Gly Gln Val
 35 40 45

Trp Val Asp Asn Asp Leu Ser Ala Thr Lys Lys Asn Val Val Arg Pro
 50 55 60

Asp Phe Glu Ala Met Ile Ala Ser Asn Pro Gln Ala Ile Val Cys Trp

		32														
	65	70	75	80												
	His	Thr	Asp	Arg	Leu	Ile	Arg	Val	Thr	Arg	Asp	Leu	Glu	Arg	Val	Ile
					85					90					95	
5	Asp	Leu	Gly	Val	Asn	Val	His	Ala	Val	Met	Ala	Gly	His	Leu	Asp	Leu
					100					105					110	
	Ser	Thr	Pro	Ala	Gly	Arg	Ala	Val	Ala	Arg	Thr	Val	Thr	Ala	Trp	Ala
10					115			120						125		
	Thr	Tyr	Glu	Gly	Glu	Gln	Lys	Ala	Glu	Arg	Gln	Lys	Leu	Ala	Asn	Ile
					130			135				140				
15	Gln	Asn	Ala	Arg	Ala	Gly	Lys	Pro	Tyr	Thr	Pro	Gly	Ile	Arg	Pro	Phe
					145			150			155			160		
	Gly	Tyr	Gly	Asp	Asp	His	Met	Thr	Ile	Val	Thr	Ala	Glu	Ala	Asp	Ala
					165				170			175				
20	Ile	Arg	Asp	Gly	Ala	Lys	Met	Ile	Leu	Asp	Gly	Trp	Ser	Leu	Ser	Ala
					180				185					190		
25	Val	Ala	Arg	Tyr	Trp	Glu	Glu	Leu	Lys	Leu	Gln	Ser	Pro	Arg	Ser	Met
					195			200				205				
	Ala	Ala	Gly	Gly	Lys	Gly	Trp	Ser	Leu	Arg	Gly	Val	Lys	Lys	Val	Leu
					210			215				220				
30	Thr	Ser	Pro	Arg	Tyr	Val	Gly	Arg	Ser	Ser	Tyr	Leu	Gly	Glu	Val	Val
					225			230			235			240		
	Gly	Asp	Ala	Gln	Trp	Pro	Pro	Ile	Leu	Asp	Pro	Asp	Val	Tyr	Tyr	Gly
					245				250				255			
35	Val	Val	Ala	Ile	Leu	Asn	Asn	Pro	Asp	Arg	Phe	Ser	Gly	Gly	Pro	Arg
					260				265				270			
40	Thr	Gly	Arg	Thr	Pro	Gly	Thr	Leu	Leu	Ala	Gly	Ile	Ala	Leu	Cys	Gly
					275			280				285				
	Glu	Cys	Gly	Lys	Thr	Val	Ser	Gly	Arg	Gly	Tyr	Arg	Gly	Val	Leu	Val
					290			295				300				
45	Tyr	Gly	Cys	Lys	Asp	Thr	His	Thr	Arg	Thr	Pro	Arg	Ser	Ile	Ala	Asp
					305			310			315			320		
	Gly	Arg	Ala	Ser	Ser	Ser	Thr	Leu	Ala	Arg	Leu	Met	Phe	Pro	Asp	Phe
					325				330				335			
50	Leu	Pro	Gly	Leu	Leu	Ala	Ser	Gly	Gln	Ala	Glu	Asp	Gly	Gln	Ser	Ala
					340				345				350			
55	Ala	Ser	Lys	His	Ser	Glu	Ala	Gln	Thr	Leu	Arg	Glu	Arg	Leu	Asp	Gly
					355				360				365			
	Leu	Ala	Thr	Ala	Tyr	Ala	Glu	Gly	Ala	Ile	Ser	Leu	Ser	Gln	Met	Thr
					370				375				380			
60	Ala	Gly	Ser	Glu	Ala	Leu	Arg	Lys	Lys	Leu	Glu	Val	Ile	Glu	Ala	Asp
					385				390			395			400	
	Leu	Val	Gly	Ser	Ala	Gly	Ile	Pro	Pro	Phe	Asp	Pro	Val	Ala	Gly	Val
					405				410				415			
65	Ala	Gly	Leu	Ile	Ser	Ser	Gly	Trp	Pro	Thr	Pro	Leu	Pro	Thr	Arg	Arg
					420				425				430			

Ala Trp Val Asp Phe Cys Leu Val Val Thr Leu Asn Thr Gln Lys Gly
 435 440 445

5 Arg His Ala Ser Ser Met Thr Val Asp Asp His Val Thr Ile Glu Trp
 450 455 460

Arg Asp Val Ala Glu
 465

10

<210> 56
 <211> 1503
 15 <212> DNA
 <213> CisA recombinase

<220>
 <221> CDS
 20 <222> (1)..(1500)

<400> 56
 gtg ata gca ata tat gta agg gta tcg acc gag gaa caa gcg atc aag 48
 Val Ile Ala Ile Tyr Val Arg Val Ser Thr Glu Glu Gln Ala Ile Lys
 25 1 5 10 15

gga tcg agc atc gac agc caa atc gag gcc tgc ata aag aaa gca ggg 96
 Gly Ser Ser Ile Asp Ser Gln Ile Glu Ala Cys Ile Lys Lys Ala Gly
 20 25 30

30

act aaa gat gtg ctg aag tat gca gat gaa gga ttt tca gga gag ctt 144
 Thr Lys Asp Val Leu Lys Tyr Ala Asp Glu Gly Phe Ser Gly Glu Leu
 35 40 45

35

tta gaa cgt ccg gct ttg aat cgc ttg agg gag gat gca agc aag gga 192
 Leu Glu Arg Pro Ala Leu Asn Arg Leu Arg Glu Asp Ala Ser Lys Gly
 50 55 60

40

ctt ata agt caa gtc att tgc gat cct gac cgt ctt tct cgg aaa 240
 Leu Ile Ser Gln Val Ile Cys Tyr Asp Pro Asp Arg Leu Ser Arg Lys
 65 70 75 80

45

tta atg aat cag cta atc att gat gac gaa ttg cga aag cga aac ata 288
 Leu Met Asn Gln Leu Ile Asp Asp Glu Leu Arg Lys Arg Asn Ile
 85 90 95

50

cct ttg att ttt gta aat ggt gaa tac gcc aat tct cca gaa ggt caa 336
 Pro Leu Ile Phe Val Asn Gly Glu Tyr Ala Asn Ser Pro Glu Gly Gln
 100 105 110

ttg ttt ttc gca atg cgc ggg gca atc tca gaa ttt gaa aaa gcc aaa 384
 Leu Phe Ala Met Arg Gly Ala Ile Ser Glu Phe Glu Lys Ala Lys
 115 120 125

55

atc aaa gaa cgg aca tca agc ggc cga ctt caa aaa atg aaa aaa ggc 432
 Ile Lys Glu Arg Thr Ser Ser Gly Arg Leu Gln Lys Met Lys Lys Gly
 130 135 140

60

atg atc att aaa gat tct aaa cta tat ggc tat aaa ttt gtt aaa gag 480
 Met Ile Ile Lys Asp Ser Lys Leu Tyr Gly Tyr Lys Phe Val Lys Glu
 145 150 155 160

65

aaa aga act ctt gag ata tta gaa gag gaa gca aaa atc att cgg atg 528
 Lys Arg Thr Leu Glu Ile Leu Glu Glu Ala Lys Ile Ile Arg Met
 165 170 175

att ttt aac tat ttc acc gat cat aaa agc cct ttt ttc ggc aga gta 576

	Ile Phe Asn Tyr Phe Thr Asp His Lys Ser Pro Phe Phe Gly Arg Val	
	180 185 190	
5	aat ggt att gct cta cat tta act cag atg ggg gtt aaa aca aaa aaa Asn Gly Ile Ala Leu His Leu Thr Gln Met Gly Val Lys Thr Lys Lys 195 200 205	624
10	ggc gcc aaa gta tgg cac agg cag gtt cggt caa ata tta atg aac Gly Ala Lys Val Trp His Arg Gln Val Val Arg Gln Ile Leu Met Asn 210 215 220	672
15	tct tcc tat aag ggt gaa cat aga cag tat aaa tat gat aca gag ggt Ser Ser Tyr Lys Gly Glu His Arg Gln Tyr Lys Tyr Asp Thr Glu Gly 225 230 235 240	720
20	tcc tat gtt tca aag cag gca ggg aac aaa tct ata att aaa ata agg Ser Tyr Val Ser Lys Gln Ala Gly Asn Lys Ser Ile Ile Lys Ile Arg 245 250 255	768
25	cct gaa gaa gaa caa atc act gtg aca att cca gca att gtt cca gct Pro Glu Glu Gln Ile Thr Val Thr Ile Pro Ala Ile Val Pro Ala 260 265 270	816
30	gaa caa tgg gat tat gct caa gaa ctc tta ggt caa agt aaa aga aaa Glu Gln Trp Asp Tyr Ala Gln Glu Leu Leu Gly Gln Ser Lys Arg Lys 275 280 285	864
35	cac ttg agt atc agc cct cac aat tac ttg tta tcg ggt ttg gtt aga His Leu Ser Ile Ser Pro His Asn Tyr Leu Leu Ser Gly Leu Val Arg 290 295 300	912
40	tgc gga aaa tgc gga aat acc atg aca ggg aag aaa aga aaa tca cat Cys Gly Lys Cys Gly Asn Thr Met Thr Gly Lys Lys Arg Lys Ser His 305 310 315 320	960
45	ggt aaa gac tac tat gta tat act tgc cgg aaa aat tat tct ggc gca Gly Lys Asp Tyr Tyr Val Tyr Thr Cys Arg Lys Asn Tyr Ser Gly Ala 325 330 335	1008
50	aag gac cgc ggc tgc gga aaa gaa atg tct gag aat aaa ttg aac cgg Lys Asp Arg Gly Cys Gly Lys Glu Met Ser Glu Asn Lys Leu Asn Arg 340 345 350	1056
55	cat gta tgg ggt gaa att ttt aaa ttc atc aca aat cct caa aag tat His Val Trp Gly Glu Ile Phe Lys Phe Ile Thr Asn Pro Gln Lys Tyr 355 360 365	1104
60	gtt tct ttt aaa gag gct gaa caa tca aat cac ctg tct gat gaa tta Val Ser Phe Lys Glu Ala Glu Gln Ser Asn His Leu Ser Asp Glu Leu 370 375 380	1152
65	gaa ctt att gaa aaa gag ata gag aaa aca aaa aaa ggc cgc aag cgt Glu Leu Ile Glu Lys Glu Ile Glu Lys Thr Lys Lys Gly Arg Lys Arg 385 390 395 400	1200
	ctt tta acg cta atc agc cta agc gat gac gat gat tta gac ata gat Leu Leu Thr Leu Ile Ser Leu Ser Asp Asp Asp Asp Leu Asp Ile Asp 405 410 415	1248
	gaa atc aaa gca caa att att gaa ctg caa aaa aag caa aat cag ctt Glu Ile Lys Ala Gln Ile Ile Glu Leu Gln Lys Lys Gln Asn Gln Leu 420 425 430	1296
	act gaa aag tgt aac aga atc cag tca aaa atg aaa gtc cta gat gat Thr Glu Lys Cys Asn Arg Ile Gln Ser Lys Met Lys Val Leu Asp Asp 435 440 445	1344

35

acg agc tca agt gaa aat gct cta aaa aga gcc atc gac tat ttt caa 1392
 Thr Ser Ser Ser Glu Asn Ala Leu Lys Arg Ala Ile Asp Tyr Phe Gln
 450 455 460

5 tca atc ggt gca gat aac tta act ctt gaa gat aaa aaa aca att gtt 1440
 Ser Ile Gly Ala Asp Asn Leu Thr Leu Glu Asp Lys Lys Thr Ile Val
 465 470 475 480

10 aac ttt atc gtg aaa gaa gtt acc att gtg gat tct gac acc ata tat 1488
 Asn Phe Ile Val Lys Glu Val Thr Ile Val Asp Ser Asp Thr Ile Tyr
 485 490 495

15 att gaa acg tat taa 1503
 Ile Glu Thr Tyr
 500

20 <210> 57
 <211> 500
 <212> PRT
 <213> CisA recombinase

25 <400> 57
 Val Ile Ala Ile Tyr Val Arg Val Ser Thr Glu Glu Gln Ala Ile Lys
 1 5 10 15

Gly Ser Ser Ile Asp Ser Gln Ile Glu Ala Cys Ile Lys Lys Ala Gly
 20 25 30

30 Thr Lys Asp Val Leu Lys Tyr Ala Asp Glu Gly Phe Ser Gly Glu Leu
 35 40 45

Leu Glu Arg Pro Ala Leu Asn Arg Leu Arg Glu Asp Ala Ser Lys Gly
 50 55 60

35 Leu Ile Ser Gln Val Ile Cys Tyr Asp Pro Asp Arg Leu Ser Arg Lys
 65 70 75 80

40 Leu Met Asn Gln Leu Ile Ile Asp Asp Glu Leu Arg Lys Arg Asn Ile
 85 90 95

Pro Leu Ile Phe Val Asn Gly Glu Tyr Ala Asn Ser Pro Glu Gly Gln
 100 105 110

45 Leu Phe Phe Ala Met Arg Gly Ala Ile Ser Glu Phe Glu Lys Ala Lys
 115 120 125

Ile Lys Glu Arg Thr Ser Ser Gly Arg Leu Gln Lys Met Lys Lys Gly
 130 135 140

50 Met Ile Ile Lys Asp Ser Lys Leu Tyr Gly Tyr Lys Phe Val Lys Glu
 145 150 155 160

55 Lys Arg Thr Leu Glu Ile Leu Glu Glu Ala Lys Ile Ile Arg Met
 165 170 175

Ile Phe Asn Tyr Phe Thr Asp His Lys Ser Pro Phe Phe Gly Arg Val
 180 185 190

60 Asn Gly Ile Ala Leu His Leu Thr Gln Met Gly Val Lys Thr Lys Lys
 195 200 205

Gly Ala Lys Val Trp His Arg Gln Val Val Arg Gln Ile Leu Met Asn
 210 215 220

65 Ser Ser Tyr Lys Gly Glu His Arg Gln Tyr Lys Tyr Asp Thr Glu Gly
 225 230 235 240

Ser Tyr Val Ser Lys Gln Ala Gly Asn Lys Ser Ile Ile Lys Ile Arg
 245 250 255

5 Pro Glu Glu Glu Gln Ile Thr Val Thr Ile Pro Ala Ile Val Pro Ala
 260 265 270

Glu Gln Trp Asp Tyr Ala Gln Glu Leu Leu Gly Gln Ser Lys Arg Lys
 275 280 285

10 His Leu Ser Ile Ser Pro His Asn Tyr Leu Leu Ser Gly Leu Val Arg
 290 295 300

15 Cys Gly Lys Cys Gly Asn Thr Met Thr Gly Lys Lys Arg Lys Ser His
 305 310 315 320

Gly Lys Asp Tyr Tyr Val Tyr Thr Cys Arg Lys Asn Tyr Ser Gly Ala
 325 330 335

20 Lys Asp Arg Gly Cys Gly Lys Glu Met Ser Glu Asn Lys Leu Asn Arg
 340 345 350

His Val Trp Gly Glu Ile Phe Lys Phe Ile Thr Asn Pro Gln Lys Tyr
 355 360 365

25 Val Ser Phe Lys Glu Ala Glu Gln Ser Asn His Leu Ser Asp Glu Leu
 370 375 380

30 Glu Leu Ile Glu Lys Glu Ile Glu Lys Thr Lys Lys Gly Arg Lys Arg
 385 390 395 400

Leu Leu Thr Leu Ile Ser Leu Ser Asp Asp Asp Asp Leu Asp Ile Asp
 405 410 415

35 Glu Ile Lys Ala Gln Ile Ile Glu Leu Gln Lys Lys Gln Asn Gln Leu
 420 425 430

Thr Glu Lys Cys Asn Arg Ile Gln Ser Lys Met Lys Val Leu Asp Asp
 435 440 445

40 Thr Ser Ser Ser Glu Asn Ala Leu Lys Arg Ala Ile Asp Tyr Phe Gln
 450 455 460

45 Ser Ile Gly Ala Asp Asn Leu Thr Leu Glu Asp Lys Lys Thr Ile Val
 465 470 475 480

Asn Phe Ile Val Lys Glu Val Thr Ile Val Asp Ser Asp Thr Ile Tyr
 485 490 495

50 Ile Glu Thr Tyr
 500

55 <210> 58
 <211> 1545
 <212> DNA
 <213> XisF recombinase

60 <220>
 <221> CDS
 <222> (1)..(1542)

65 <400> 58
 atg gaa aat tgg ggt tac gcg aga gtc agc ggt gag gaa cag caa aca 48
 Met Glu Asn Trp Gly Tyr Ala Arg Val Ser Gly Glu Glu Gln Gln Thr
 1 5 10 15

5	gat aaa ggt gcg ttg cgt aaa caa ata gaa cgc ttg cgt aat gct gga Asp Lys Gly Ala Leu Arg Lys Gln Ile Glu Arg Leu Arg Asn Ala Gly 20 25 30	96
10	tgt tca aaa gtg tac tgg gat att caa tcg cgg aca act gaa gtc aga Cys Ser Lys Val Tyr Trp Asp Ile Gln Ser Arg Thr Thr Glu Val Arg 35 40 45	144
15	gaa ggg cta caa caa tta att aat gac tta aag aca tct tca aca ggt Glu Gly Leu Gln Gln Leu Ile Asn Asp Leu Lys Thr Ser Ser Thr Gly 50 55 60	192
20	aag gta aaa tca ctg caa ttt acc cgc att gat cgc atc ggc tca tca Lys Val Lys Ser Ile Gln Phe Thr Arg Ile Asp Arg Ile Gly Ser Ser 65 70 75 80	240
25	tcg cgg ttg ttt tat tca ttg tta gag gta tta cgt tcc aag gga att Ser Arg Leu Phe Tyr Ser Leu Leu Glu Val Leu Arg Ser Lys Gly Ile 85 90 95	288
30	aaa ctg ata gcc tta gat caa ggc gtt gac cca gac agc ctt ggc ggg Lys Leu Ile Ala Leu Asp Gln Gly Val Asp Pro Asp Ser Leu Gly Gly 100 105 110	336
35	gaa cta aca att gat atg tta ctg gct gct gcc aaa ttt gag gta aga Glu Leu Thr Ile Asp Met Leu Ala Ala Lys Phe Glu Val Arg 115 120 125	384
40	atg gtg acg gag agg tta aaa agc gaa cgt cgt cat agg gtg aac caa Met Val Thr Glu Arg Leu Lys Ser Glu Arg Arg His Arg Val Asn Gln 130 135 140	432
45	gga aaa agt cac cga gtt gcc cca tta gga tac cgc aaa gat aaa gat Gly Lys Ser His Arg Val Ala Pro Leu Gly Tyr Arg Lys Asp Lys Asp 145 150 155 160	480
50	aaa tat ata cgc gat cgc tca cca tgt gtt tgc tta cta gaa gga cgc Lys Tyr Ile Arg Asp Arg Ser Pro Cys Val Cys Leu Leu Glu Gly Arg 165 170 175	528
55	aga gaa tta acg gtg tct gac tta gcc cag tat att ttt cac act ttt Arg Glu Leu Thr Val Ser Asp Leu Ala Gln Tyr Ile Phe His Thr Phe 180 185 190	576
60	ttt gag tgc ggt tcc gtt gct gct act gtg cgt aag ctg cac tca gat Phe Glu Cys Gly Ser Val Ala Ala Thr Val Arg Lys Leu His Ser Asp 195 200 205	624
65	ttt ggt ata gaa aca aaa gtt ctg aat tgg aac aag cta gaa aaa tct Phe Gly Ile Glu Thr Lys Val Leu Asn Trp Asn Lys Leu Glu Lys Ser 210 215 220	672
70	tcc cgg att gtt ggc gac gac gac tta gat aaa att gca ttt aca cca Ser Arg Ile Val Gly Asp Asp Asp Leu Asp Lys Ile Ala Phe Thr Pro 225 230 235 240	720
75	aat aaa act aac cac ccc ttg cgt tat ccc tgg tct ggg cta aga tgg Asn Lys Thr Asn His Pro Leu Arg Tyr Pro Trp Ser Gly Leu Arg Trp 245 250 255	768
80	tca atc cct ggt tta aaa gcg tta tta gtt aac cct gtt tac gcc ggg Ser Ile Pro Gly Leu Lys Ala Leu Leu Val Asn Pro Val Tyr Ala Gly 260 265 270	816
85	ggt ttg ccc ttt gat act tac gtt aaa tca aaa gga aaa cgc aag cat Gly Leu Pro Phe Asp Thr Tyr Val Lys Ser Lys Gly Lys Arg Lys His	864

	275	280	285	
5	ttt gac gag tgg aaa gta aaa tgg gga acc cac gac gat gag gca atc Phe Asp Glu Trp Lys Val Lys Trp Gly Thr His Asp Asp Glu Ala Ile 290 295 300			912
10	att acc tgt gag gaa cat gaa aga ata aaa cag atg att cga gac aat Ile Thr Cys Glu Glu His Glu Arg Ile Lys Gln Met Ile Arg Asp Asn 305 310 315 320			960
15	cgc aat aat cga tgg gct gca aga gaa aac gaa gta aac cca ttt Arg Asn Asn Arg Trp Ala Ala Arg Glu Glu Asn Glu Val Asn Pro Phe 325 330 335			1008
20	tct aat tta ctt aaa tgt acc cat tgc ggc ggc tca atg aca cgc cac Ser Asn Leu Leu Lys Cys Thr His Cys Gly Ser Met Thr Arg His 340 345 350			1056
25	gcc aaa cgt gta gat aag agt gga caa gct atc tat tat tat cag tgc Ala Lys Arg Val Asp Lys Ser Gly Gln Ala Ile Tyr Tyr Gln Cys 355 360 365			1104
30	cga ttg tat aaa gct ggc aac tgt agc aat aaa aat atg att tca tcc Arg Leu Tyr Lys Ala Gly Asn Cys Ser Asn Lys Asn Met Ile Ser Ser 370 375 380			1152
35	aaa ata tta gat atc caa gta atg gat tta ttg gca caa gaa gcc gaa Lys Ile Leu Asp Ile Gln Val Met Asp Leu Leu Ala Gln Glu Ala Glu 385 390 395 400			1200
40	cgt tta gca aat ttg gtg gaa aca gat gag ccg ctt att gta gaa gaa Arg Leu Ala Asn Leu Val Glu Thr Asp Glu Pro Leu Ile Val Glu Glu 405 410 415			1248
45	ccc cca gaa gta aaa acg ctg cgc gca tcc ctg aat agt ctg gaa aca Pro Pro Glu Val Lys Thr Leu Arg Ala Ser Leu Asn Ser Leu Glu Thr 420 425 430			1296
50	ttg cca gca agt tca gca att gaa caa att aaa aat gac ctc aaa gaa Leu Pro Ala Ser Ser Ala Ile Glu Gln Ile Lys Asn Asp Leu Lys Glu 435 440 445			1344
55	cag att gcg atc gca cta gga gca acc aat aat gct tct aaa caa tct Gln Ile Ala Ile Ala Leu Gly Ala Thr Asn Asn Ala Ser Lys Gln Ser 450 455 460			1392
60	ctg att gcc aag gaa aga att ata caa gct ttt gct cat aaa agt tac Leu Ile Ala Lys Glu Arg Ile Ile Gln Ala Phe Ala His Lys Ser Tyr 465 470 475 480			1440
65	tgg caa gga cta aac gct caa gat aaa cga gca atc ctc aat ggt tgc Trp Gln Gly Leu Asn Ala Gln Asp Lys Arg Ala Ile Leu Asn Gly Cys 485 490 495			1488
	gta aaa aaa atc tcc gta gat ggt aac ttt gtt aca gct att gag tat Val Lys Lys Ile Ser Val Asp Gly Asn Phe Val Thr Ala Ile Glu Tyr 500 505 510			1536
	cgt tac tag Arg Tyr			1545

<210> 59

<211> 514

<212> PRT

<213> XisF recombinase

<400> 59
 Met Glu Asn Trp Gly Tyr Ala Arg Val Ser Gly Glu Glu Gln Gln Thr
 1 5 10 15

5 Asp Lys Gly Ala Leu Arg Lys Gln Ile Glu Arg Leu Arg Asn Ala Gly
 20 25 30

Cys Ser Lys Val Tyr Trp Asp Ile Gln Ser Arg Thr Thr Glu Val Arg
 35 40 45

10 Glu Gly Leu Gln Gln Leu Ile Asn Asp Leu Lys Thr Ser Ser Thr Gly
 50 55 60

15 Lys Val Lys Ser Leu Gln Phe Thr Arg Ile Asp Arg Ile Gly Ser Ser
 65 70 75 80

Ser Arg Leu Phe Tyr Ser Leu Leu Glu Val Leu Arg Ser Lys Gly Ile
 85 90 95

20 Lys Leu Ile Ala Leu Asp Gln Gly Val Asp Pro Asp Ser Leu Gly Gly
 100 105 110

Glu Leu Thr Ile Asp Met Leu Leu Ala Ala Ala Lys Phe Glu Val Arg
 115 120 125

25 Met Val Thr Glu Arg Leu Lys Ser Glu Arg Arg His Arg Val Asn Gln
 130 135 140

Gly Lys Ser His Arg Val Ala Pro Leu Gly Tyr Arg Lys Asp Lys Asp
 30 145 150 155 160

Lys Tyr Ile Arg Asp Arg Ser Pro Cys Val Cys Leu Leu Glu Gly Arg
 165 170 175

35 Arg Glu Leu Thr Val Ser Asp Leu Ala Gln Tyr Ile Phe His Thr Phe
 180 185 190

Phe Glu Cys Gly Ser Val Ala Ala Thr Val Arg Lys Leu His Ser Asp
 195 200 205

40 Phe Gly Ile Glu Thr Lys Val Leu Asn Trp Asn Lys Leu Glu Lys Ser
 210 215 220

Ser Arg Ile Val Gly Asp Asp Asp Leu Asp Lys Ile Ala Phe Thr Pro
 45 225 230 235 240

Asn Lys Thr Asn His Pro Leu Arg Tyr Pro Trp Ser Gly Leu Arg Trp
 245 250 255

50 Ser Ile Pro Gly Leu Lys Ala Leu Leu Val Asn Pro Val Tyr Ala Gly
 260 265 270

Gly Leu Pro Phe Asp Thr Tyr Val Lys Ser Lys Gly Lys Arg Lys His
 275 280 285

55 Phe Asp Glu Trp Lys Val Lys Trp Gly Thr His Asp Asp Glu Ala Ile
 290 295 300

Ile Thr Cys Glu Glu His Glu Arg Ile Lys Gln Met Ile Arg Asp Asn
 60 305 310 315 320

Arg Asn Asn Arg Trp Ala Ala Arg Glu Glu Asn Glu Val Asn Pro Phe
 325 330 335

65 Ser Asn Leu Leu Lys Cys Thr His Cys Gly Gly Ser Met Thr Arg His
 340 345 350

40

Ala Lys Arg Val Asp Lys Ser Gly Gln Ala Ile Tyr Tyr Tyr Gln Cys
 355 360 365

5 Arg Leu Tyr Lys Ala Gly Asn Cys Ser Asn Lys Asn Met Ile Ser Ser
 370 375 380

Lys Ile Leu Asp Ile Gln Val Met Asp Leu Leu Ala Gln Glu Ala Glu
 385 390 395 400

10 Arg Leu Ala Asn Leu Val Glu Thr Asp Glu Pro Leu Ile Val Glu Glu
 405 410 415

Pro Pro Glu Val Lys Thr Leu Arg Ala Ser Leu Asn Ser Leu Glu Thr
 420 425 430

15 Leu Pro Ala Ser Ser Ala Ile Glu Gln Ile Lys Asn Asp Leu Lys Glu
 435 440 445

20 Gln Ile Ala Ile Ala Leu Gly Ala Thr Asn Asn Ala Ser Lys Gln Ser
 450 455 460

Leu Ile Ala Lys Glu Arg Ile Ile Gln Ala Phe Ala His Lys Ser Tyr
 465 470 475 480

25 Trp Gln Gly Leu Asn Ala Gln Asp Lys Arg Ala Ile Leu Asn Gly Cys
 485 490 495

Val Lys Lys Ile Ser Val Asp Gly Asn Phe Val Thr Ala Ile Glu Tyr
 500 505 510

30 Arg Tyr

35 <210> 60

<211> 2124

<212> DNA

40 <213> Transposon Tn4451

<220>

<221> CDS

45 <222> (1)...(2121)

45 <400> 60

atg tca agg act tca aga att aca gca ctt tac gag cgt ttg tca aga 48
 Met Ser Arg Thr Ser Arg Ile Thr Ala Leu Tyr Glu Arg Leu Ser Arg
 1 5 10 15

50 gat gat gac ctt act ggc gag aat tct att acc aat caa aag aaa 96
 Asp Asp Asp Leu Thr Gly Glu Ser Asn Ser Ile Thr Asn Gln Lys Lys
 20 25 30

55 tac ctc gaa gat tat gcc cgt agg aat ggt ttt gag aac att cgc cat 144
 Tyr Leu Glu Asp Tyr Ala Arg Arg Asn Gly Phe Glu Asn Ile Arg His
 35 40 45

60 ttt acc gat gac gga ttt tcg ggt gta aat ttc aat cgc cct ggc ttt 192
 Phe Thr Asp Asp Gly Phe Ser Gly Val Asn Phe Asn Arg Pro Gly Phe
 50 55 60

65 caa tct ctg ata aaa gaa gtt gaa gca gga aat gta gaa acc ttg att 240
 Gln Ser Leu Ile Lys Glu Val Glu Ala Gly Asn Val Glu Thr Leu Ile
 65 70 75 80

65 gtt aag gat atg agc cga ttg ggg cga aat tat ctg caa gta ggt ttt 288
 Val Lys Asp Met Ser Arg Leu Gly Arg Asn Tyr Leu Gln Val Gly Phe

	85	90	95	
		41		
5	tat acg gaa gtt ctg ttt cca cag aaa aat gtc cgt ttc ctt gca att Tyr Thr Glu Val Leu Phe Pro Gln Lys Asn Val Arg Phe Leu Ala Ile 100 105 110			336
	aac aac agt att gac agt aac aac gct tcg gat aat gac ttt gct ccg Asn Asn Ser Ile Asp Ser Asn Asn Ala Ser Asp Asn Asp Phe Ala Pro 115 120 125			384
10	ttt ttg aat att atg aac gaa tgg tat gcc aaa gac aca agc aac aaa Phe Leu Asn Ile Met Asn Glu Trp Tyr Ala Lys Asp Thr Ser Asn Lys 130 135 140			432
15	atc aag gct ata ttc gat gcc cgt atg aaa gac gga aag cgt tgt agc Ile Lys Ala Ile Phe Asp Ala Arg Met Lys Asp Gly Lys Arg Cys Ser 145 150 155 160			480
20	ggt tct atc cct tat ggg tat aac cga ctg ccg agc gac aaa caa acg Gly Ser Ile Pro Tyr Gly Tyr Asn Arg Leu Pro Ser Asp Lys Gln Thr 165 170 175			528
25	ctt gtg gtt gac cct gtg gct tcg gaa gtg gta aag cgt atc ttt act Leu Val Val Asp Pro Val Ala Ser Glu Val Val Lys Arg Ile Phe Thr 180 185 190			576
30	ctt gcc aat gat ggc aaa agt aca agg gca atc gca gaa ata ctg acc Leu Ala Asn Asp Gly Lys Ser Thr Arg Ala Ile Ala Glu Ile Leu Thr 195 200 205			624
	gaa gaa aaa gtt tta acc cct gcg gca tac gca aag gaa tac cac ccc Glu Glu Lys Val Leu Thr Pro Ala Ala Tyr Ala Lys Glu Tyr His Pro 210 215 220			672
35	gaa cag tac aac ggc aac aag ttc aca aac cct tat ctt tgg gca atg Glu Gln Tyr Asn Gly Asn Lys Phe Thr Asn Pro Tyr Leu Trp Ala Met 225 230 235 240			720
40	tca acg ata aga aat att tta ggc agg cag gaa tat ctc ggt cac acc Ser Thr Ile Arg Asn Ile Leu Gly Arg Gln Glu Tyr Leu Gly His Thr 245 250 255			768
45	gtt ttg cga aag tcg gta agc aca aat ttc aaa ctt cac aag aga aaa Val Leu Arg Lys Ser Val Ser Thr Asn Phe Lys Leu His Lys Arg Lys 260 265 270			816
	agc aca gac gaa gaa cag tat gta ttt ccg aat aca cac gag cct Ser Thr Asp Glu Glu Gln Tyr Val Phe Pro Asn Thr His Glu Pro 275 280 285			864
50	atc ata tcg cag gaa ctt tgg gac agc gtt caa aaa cgc aga agc aga Ile Ile Ser Gln Glu Leu Trp Asp Ser Val Gln Lys Arg Arg Ser Arg 290 295 300			912
55	gta aat cgt gcc tcg gct tgg gga acg cac agc aac cgt tta agc gga Val Asn Arg Ala Ser Ala Trp Gly Thr His Ser Asn Arg Leu Ser Gly 305 310 315 320			960
60	tat ttg tac tgt gcc gat tgc gga aga aga atg act ttg cag aca cat Tyr Leu Tyr Cys Ala Asp Cys Gly Arg Arg Met Thr Leu Gln Thr His 325 330 335			1008
65	tac agc aaa aaa gac ggt tct tcg cag tat tct tac cgt tgc ggt ggg Tyr Ser Lys Lys Asp Gly Ser Val Gln Tyr Ser Tyr Arg Cys Gly Gly 340 345 350			1056
	tat gca agc aga gtg aac agt tgt acc agt cat tcg att agt acc gat			1104

	Tyr Ala Ser Arg Val Asn Ser Cys Thr Ser His Ser Ile Ser Thr Asp			
	355	360	365	
5	aat gtt gaa gcc ttg ata tta tca tct gtc aaa cgc ttt tca agg ttt	1152		
	Asn Val Glu Ala Leu Ile Leu Ser Ser Val Lys Arg Phe Ser Arg Phe			
	370	375	380	
10	gtt ctg aat gat gaa caa gca ttt gct ttg gaa ctg caa tct ctt tgg	1200		
	Val Leu Asn Asp Glu Gln Ala Phe Ala Leu Glu Leu Gln Ser Leu Trp			
	385	390	395	400
	aat gaa aaa cag gag gaa aag ccg aaa cac aat caa tcg gaa ctg caa	1248		
	Asn Glu Lys Gln Glu Lys Pro Lys His Asn Gln Ser Glu Leu Gln			
	405	410	415	
15	cgc tgt cag aaa cgc tat gac gaa ctc tct acc ctt gtt cgt ggc ttg	1296		
	Arg Cys Gln Lys Arg Tyr Asp Glu Leu Ser Thr Leu Val Arg Gly Leu			
	420	425	430	
20	tat gaa aat ctt atg tcg gga tta ctg ccc gaa aga cag tat aag caa	1344		
	Tyr Glu Asn Leu Met Ser Gly Leu Leu Pro Glu Arg Gln Tyr Lys Gln			
	435	440	445	
25	ctg atg aaa cag tat gat gac gag cag gca gag ttg gaa acg aaa atg	1392		
	Leu Met Lys Gln Tyr Asp Asp Glu Gln Ala Glu Leu Glu Thr Lys Met			
	450	455	460	
30	gaa acg atg aaa aca gaa ctt gcc gaa gaa aaa gta agt tcc gtt gat	1440		
	Glu Thr Met Lys Thr Glu Leu Ala Glu Glu Lys Val Ser Ser Val Asp			
	465	470	475	480
	att aag cat ttc att tog ctg ata cgc aag tgt aaa aat cct acg gaa	1488		
	Ile Lys His Phe Ile Ser Leu Ile Arg Lys Cys Lys Asn Pro Thr Glu			
	485	490	495	
35	atc tcc gat aca atg ttt aat gaa ctt gtt gat aag ata gtg gtt tat	1536		
	Ile Ser Asp Thr Met Phe Asn Glu Leu Val Asp Lys Ile Val Val Tyr			
	500	505	510	
40	gaa gca gag ggt gtg gga aaa gca cga aca caa aag gtc gat att tat	1584		
	Glu Ala Glu Gly Val Gly Lys Ala Arg Thr Gln Lys Val Asp Ile Tyr			
	515	520	525	
45	ttt aac tat gtc ggt caa gtg gat att gcc tat acc gaa gaa gaa ctt	1632		
	Phe Asn Tyr Val Gly Gln Val Asp Ile Ala Tyr Thr Glu Glu Glu Leu			
	530	535	540	
50	gcc gag ata gaa aca cag aaa gag cag gag gaa cag caa cgc ttg gca	1680		
	Ala Glu Ile Glu Thr Gln Lys Glu Gln Glu Glu Gln Gln Arg Leu Ala			
	545	550	555	560
	aga cag cgc aag cgt gaa aaa gcc tac cga gaa aag cga aag gca cag	1728		
	Arg Gln Arg Lys Arg Glu Lys Ala Tyr Arg Glu Lys Arg Lys Ala Gln			
	565	570	575	
55	aaa atc gct gaa aac ggt ggc gaa atc gtt aag aca aag gtt tgc cct	1776		
	Lys Ile Ala Glu Asn Gly Gly Glu Ile Val Lys Thr Lys Val Cys Pro			
	580	585	590	
60	cat tgc aac aaa gag ttt atc ccg aca agc aac cga cag gtg ttc tgt	1824		
	His Cys Asn Lys Glu Phe Ile Pro Thr Ser Asn Arg Gln Val Phe Cys			
	595	600	605	
65	tcc aaa gag tgc tgc tat caa gca agg caa gac aaa aag aaa aca gac	1872		
	Ser Lys Glu Cys Cys Tyr Gln Ala Arg Gln Asp Lys Lys Lys Thr Asp			
	610	615	620	

43

cga gaa gca gaa cga gga aat cac tat tac cga cag cgt gta tgt gct 1920
 Arg Glu Ala Glu Arg Gly Asn His Tyr Tyr Arg Gln Arg Val Cys Ala
 625 630 635 640

5 gtg tgc ggc aat tcc tat tgg cct aca cac agc caa cag aaa ttc tgc 1968
 Val Cys Gly Asn Ser Tyr Trp Pro Thr His Ser Gln Gln Lys Phe Cys
 645 650 655

10 tcc gaa gaa tgt caa agg gta aat cac aat aag aaa aca ttg gaa ttt 2016
 Ser Glu Glu Cys Gln Arg Val Asn His Asn Lys Lys Thr Leu Glu Phe
 660 665 670

15 tac cac cat aaa aaa gaa aag gag aag ctg caa tgc aaa gat tta tca 2064
 Tyr His Lys Lys Glu Lys Leu Gln Cys Lys Asp Leu Ser
 675 680 685

20 cag acg aaa gaa cgg gta tcc gat atg aac tta tcg ggg act att act 2112
 Gln Thr Lys Glu Arg Val Ser Asp Met Asn Leu Ser Gly Thr Ile Thr
 690 695 700

20 acc cct gct taa 2124
 Thr Pro Ala
 705

25 <210> 61
 <211> 707
 <212> PRT
 30 <213> Transposon Tn4451

35 <400> 61
 Met Ser Arg Thr Ser Arg Ile Thr Ala Leu Tyr Glu Arg Leu Ser Arg
 1 5 10 15

40 Asp Asp Asp Leu Thr Gly Glu Ser Asn Ser Ile Thr Asn Gln Lys Lys
 20 25 30

45 Tyr Leu Glu Asp Tyr Ala Arg Arg Asn Gly Phe Glu Asn Ile Arg His
 35 40 45

50 Phe Thr Asp Asp Gly Phe Ser Gly Val Asn Phe Asn Arg Pro Gly Phe
 50 55 60

55 Gln Ser Leu Ile Lys Glu Val Glu Ala Gly Asn Val Glu Thr Leu Ile
 65 70 75 80

60 Val Lys Asp Met Ser Arg Leu Gly Arg Asn Tyr Leu Gln Val Gly Phe
 85 90 95

65 Tyr Thr Glu Val Leu Phe Pro Gln Lys Asn Val Arg Phe Leu Ala Ile
 100 105 110

70 Asn Asn Ser Ile Asp Ser Asn Asn Ala Ser Asp Asn Asp Phe Ala Pro
 115 120 125

75 Phe Leu Asn Ile Met Asn Glu Trp Tyr Ala Lys Asp Thr Ser Asn Lys
 130 135 140

80 Ile Lys Ala Ile Phe Asp Ala Arg Met Lys Asp Gly Lys Arg Cys Ser
 145 150 155 160

85 Gly Ser Ile Pro Tyr Gly Tyr Asn Arg Leu Pro Ser Asp Lys Gln Thr
 165 170 175

90 Leu Val Val Asp Pro Val Ala Ser Glu Val Val Lys Arg Ile Phe Thr
 180 185 190

44

Leu Ala Asn Asp Gly Lys Ser Thr Arg Ala Ile Ala Glu Ile Leu Thr
 195 200 205

5 Glu Glu Lys Val Leu Thr Pro Ala Ala Tyr Ala Lys Glu Tyr His Pro
 210 215 220

Glu Gln Tyr Asn Gly Asn Lys Phe Thr Asn Pro Tyr Leu Trp Ala Met
 225 230 235 240

10 Ser Thr Ile Arg Asn Ile Leu Gly Arg Gln Glu Tyr Leu Gly His Thr
 245 250 255

Val Leu Arg Lys Ser Val Ser Thr Asn Phe Lys Leu His Lys Arg Lys
 15 260 265 270

Ser Thr Asp Glu Glu Glu Gln Tyr Val Phe Pro Asn Thr His Glu Pro
 275 280 285

20 Ile Ile Ser Gln Glu Leu Trp Asp Ser Val Gln Lys Arg Arg Ser Arg
 290 295 300

Val Asn Arg Ala Ser Ala Trp Gly Thr His Ser Asn Arg Leu Ser Gly
 305 310 315 320

25 Tyr Leu Tyr Cys Ala Asp Cys Gly Arg Arg Met Thr Leu Gln Thr His
 325 330 335

Tyr Ser Lys Lys Asp Gly Ser Val Gln Tyr Ser Tyr Arg Cys Gly Gly
 340 345 350

30 Tyr Ala Ser Arg Val Asn Ser Cys Thr Ser His Ser Ile Ser Thr Asp
 355 360 365

Asn Val Glu Ala Leu Ile Leu Ser Ser Val Lys Arg Phe Ser Arg Phe
 35 370 375 380

Val Leu Asn Asp Glu Gln Ala Phe Ala Leu Glu Leu Gln Ser Leu Trp
 385 390 395 400

40 Asn Glu Lys Gln Glu Lys Pro Lys His Asn Gln Ser Glu Leu Gln
 405 410 415

Arg Cys Gln Lys Arg Tyr Asp Glu Leu Ser Thr Leu Val Arg Gly Leu
 420 425 430

45 Tyr Glu Asn Leu Met Ser Gly Leu Leu Pro Glu Arg Gln Tyr Lys Gln
 435 440 445

Leu Met Lys Gln Tyr Asp Asp Glu Gln Ala Glu Leu Glu Thr Lys Met
 50 450 455 460

Glu Thr Met Lys Thr Glu Leu Ala Glu Glu Lys Val Ser Ser Val Asp
 465 470 475 480

55 Ile Lys His Phe Ile Ser Leu Ile Arg Lys Cys Lys Asn Pro Thr Glu
 485 490 495

Ile Ser Asp Thr Met Phe Asn Glu Leu Val Asp Lys Ile Val Val Tyr
 60 500 505 510

Glu Ala Glu Gly Val Gly Lys Ala Arg Thr Gln Lys Val Asp Ile Tyr
 515 520 525

Phe Asn Tyr Val Gly Gin Val Asp Ile Ala Tyr Thr Glu Glu Glu Leu
 65 530 535 540

Ala Glu Ile Glu Thr Gln Lys Glu Gln Glu Gln Arg Leu Ala

		45														
	545	550	555	560												
	Arg Gln Arg Lys Arg Glu Lys Ala Tyr Arg Glu Lys Arg Lys Ala Gln															
	565	570	575													
5	Lys Ile Ala Glu Asn Gly Gly Glu Ile Val Lys Thr Lys Val Cys Pro															
	580	585	590													
10	His Cys Asn Lys Glu Phe Ile Pro Thr Ser Asn Arg Gln Val Phe Cys															
	595	600	605													
	Ser Lys Glu Cys Cys Tyr Gln Ala Arg Gln Asp Lys Lys Lys Thr Asp															
	610	615	620													
15	Arg Glu Ala Glu Arg Gly Asn His Tyr Tyr Arg Gln Arg Val Cys Ala															
	625	630	635	640												
	Val Cys Gly Asn Ser Tyr Trp Pro Thr His Ser Gln Gln Lys Phe Cys															
20	645	650	655													
	Ser Glu Glu Cys Gln Arg Val Asn His Asn Lys Lys Thr Leu Glu Phe															
	660	665	670													
25	Tyr His His Lys Lys Glu Lys Glu Lys Leu Gln Cys Lys Asp Leu Ser															
	675	680	685													
	Gln Thr Lys Glu Arg Val Ser Asp Met Asn Leu Ser Gly Thr Ile Thr															
	690	695	700													
30	Thr Pro Ala															
	705															
35	<210> 62															
	<211> 1420															
	<212> DNA															
	<213> XisA recombinase															
40	<220>															
	<221> CDS															
	<222> (1)...(1416)															
45	<400> 62															
	atg caa aat cag ggt caa gac aaa tat caa caa gcc ttt gca gac tta															
	Met	Gln	Asn	Gln	Gly	Gln	Asp	Lys	Tyr	Gln	Ala	Phe	Ala	Asp	Leu	
	1	5	10	15												
50	gag cca ott tca tct acc gac ggc agt ttt ctc ggc tca agt ctg caa															
	Glu	Pro	Leu	Ser	Ser	Thr	Asp	Gly	Ser	Phe	Leu	Gly	Ser	Ser	Leu	Gln
	20	25	30													
55	gca cag cag caa aga gaa cac atg aga aca aaa gta cta caa gac cta															
	Ala	Gln	Gln	Arg	Glu	His	Met	Arg	Thr	Lys	Val	Leu	Gln	Asp	Leu	
	35	40	45													
60	gac aag gta aat ctg cgt ttg aag tct gca aag acg aaa gtc tca gtt															
	Asp	Lys	Val	Asn	Leu	Arg	Leu	Lys	Ser	Ala	Lys	Thr	Lys	Val	Ser	Val
	50	55	60													
65	cga gaa tct aac gga agt ctg caa tta cga gca acg tta cca att aaa															
	Arg	Glu	Ser	Asn	Gly	Ser	Leu	Gln	Leu	Arg	Ala	Thr	Leu	Pro	Ile	Lys
	65	70	75	80												
70	cct gga gat aag gac acc aac ggt aca ggc aga aag caa tac aat ctc															
	Pro	Gly	Asp	Lys	Asp	Thr	Asn	Gly	Thr	Gly	Arg	Lys	Gln	Tyr	Asn	Leu
	85	90	95													

5	agc ttg aat atc cct gca aac ttg gat gga ctg aag acg gct gag gaa Ser Leu Asn Ile Pro Ala Asn Leu Asp Gly Leu Lys Thr Ala Glu Glu 100 105 110	336
10	gaa gct tat gaa tta ggt aaa tta atc gct cgg aaa acc ttt gaa tgg Glu Ala Tyr Glu Leu Gly Lys Leu Ile Ala Arg Lys Thr Phe Glu Trp 115 120 125	384
15	aat gat aaa tat tta ggc aaa gaa gcc act aaa aaa gat tca caa aca Asn Asp Lys Tyr Leu Gly Lys Glu Ala Thr Lys Lys Asp Ser Gln Thr 130 135 140	432
20	ata ggt gat tta cta gaa aaa ttt gca gaa gag tat ttt aaa acc cat Ile Gly Asp Leu Leu Glu Lys Phe Ala Glu Glu Tyr Phe Lys Thr His 145 150 155 160	480
25	aaa cgc acc act aaa agc gaa cat acc ttt ttt tac tat ttt tcc cgc Lys Arg Thr Thr Lys Ser Glu His Thr Phe Phe Tyr Tyr Phe Ser Arg 165 170 175	528
30	acc caa cga tat acc aat tcc aaa gat tta gca acg gcg gaa aat ctc Thr Gln Arg Tyr Thr Asn Ser Lys Asp Leu Ala Thr Ala Glu Asn Leu 180 185 190	576
35	atc aat tca att gag caa atc gat aaa gaa tgg gcg aga tat aat gcc Ile Asn Ser Ile Glu Gln Ile Asp Lys Glu Trp Ala Arg Tyr Asn Ala 195 200 205	624
40	gcc aga gcc ata tca gct ttt tgc ata aca ttc aat ata gaa att gat Ala Arg Ala Ile Ser Ala Phe Cys Ile Thr Phe Asn Ile Glu Ile Asp 210 215 220	672
45	ttg tcc cag tat tcc aaa atg cct gat cgc aat tcg cgc aac atc ccc Leu Ser Gln Tyr Ser Lys Met Pro Asp Arg Asn Ser Arg Asn Ile Pro 225 230 235 240	720
50	aca gat gca gaa ata cta tca gga att acc aaa ttt gaa gac tat cta Thr Asp Ala Glu Ile Leu Ser Gly Ile Thr Lys Phe Glu Asp Tyr Leu 245 250 255	768
55	gtt acc aga gga aat caa gtt aat gaa gat gta aaa gat agc tgg caa Val Thr Arg Gly Asn Gln Val Asn Glu Asp Val Lys Asp Ser Trp Gln 260 265 270	816
60	ctt tgg cgc tgg aca tat gga atg tta gca gtt ttt ggt tta cgc ccc Leu Trp Arg Trp Thr Tyr Gly Met Leu Ala Val Phe Gly Leu Arg Pro 275 280 285	864
65	agg gaa att ttt att aac cct aat att gat tgg tgg tta agc aaa gag Arg Glu Ile Phe Ile Asn Pro Asn Ile Asp Trp Trp Leu Ser Lys Glu 290 295 300	912
	aat ata gac ctc aca tgg aaa gta gac aaa gaa tgt aaa act ggt gaa Asn Ile Asp Leu Thr Trp Lys Val Asp Lys Glu Cys Lys Thr Gly Glu 305 310 315 320	960
	aga caa gca tta ccc tta cat aaa gaa tgg att gat gag ttt gat tta Arg Gln Ala Leu Pro Leu His Lys Glu Trp Ile Asp Glu Phe Asp Leu 325 330 335	1008
	aga aat ccg aaa tat tta gaa atg ctg gca aca gca att agt aaa aaa Arg Asn Pro Lys Tyr Leu Glu Met Leu Ala Thr Ala Ile Ser Lys Lys 340 345 350	1056
	gat aaa aca aat cat gct gaa ata aca gcc tta act cag cgt att agt Asp Lys Thr Asn His Ala Glu Ile Thr Ala Leu Thr Gln Arg Ile Ser	1104

47

	355	360	365	
5	tgg tgg ttt cgg aaa gtc gaa tta gat ttt aaa ccc tat gat tta cgt Trp Trp Phe Arg Lys Val Glu Leu Asp Phe Lys Pro Tyr Asp Leu Arg 370 375 380			1152
10	cac gcc tgg gca atc aga gcg cat att tta ggc ata cca atc aaa gcg His Ala Trp Ala Ile Arg Ala His Ile Leu Gly Ile Pro Ile Lys Ala 385 390 395 400			1200
15	gcg gct gat aat ttg ggg cat agt atg cag gtt cat aca caa acc tat Ala Ala Asp Asn Leu Gly His Ser Met Gln Val His Thr Gln Thr Tyr 405 410 415			1248
20	cag cgc tgg ttc tcg cta gat atg cgg aag tta gcg att aat cag gct Gln Arg Trp Phe Ser Leu Asp Met Arg Lys Leu Ala Ile Asn Gln Ala 420 425 430			1296
25	ttg act aag agg aat gaa ttt gag gtg att agg gag gag aat gct aaa Leu Thr Lys Arg Asn Glu Phe Glu Val Ile Arg Glu Glu Asn Ala Lys 435 440 445			1344
30	ttg cag ata gaa aat gaa agg ttg agg atg gaa att gag aag tta aag Leu Gln Ile Glu Asn Glu Arg Leu Arg Met Glu Ile Glu Lys Leu Lys 450 455 460			1392
35	atg gaa ata gct tat aag aat agt tgag Met Glu Ile Ala Tyr Lys Asn Ser 465 470			1420
40	<210> 63 <211> 472 <212> PRT <213> XisA recombinase			
45	<400> 63 Met Gln Asn Gln Gly Gln Asp Lys Tyr Gln Gln Ala Phe Ala Asp Leu 1 5 10 15			
50	Glu Pro Leu Ser Ser Thr Asp Gly Ser Phe Leu Gly Ser Ser Leu Gln 20 25 30			
55	Ala Gln Gln Gln Arg Glu His Met Arg Thr Lys Val Leu Gln Asp Leu 35 40 45			
60	Asp Lys Val Asn Leu Arg Leu Lys Ser Ala Lys Thr Lys Val Ser Val 50 55 60			
65	Arg Glu Ser Asn Gly Ser Leu Gln Leu Arg Ala Thr Leu Pro Ile Lys 65 70 75 80			
70	Pro Gly Asp Lys Asp Thr Asn Gly Thr Gly Arg Lys Gln Tyr Asn Leu 85 90 95			
75	Ser Leu Asn Ile Pro Ala Asn Leu Asp Gly Leu Lys Thr Ala Glu Glu 100 105 110			
80	Glu Ala Tyr Glu Leu Gly Lys Leu Ile Ala Arg Lys Thr Phe Glu Trp 115 120 125			
85	Asn Asp Lys Tyr Leu Gly Lys Glu Ala Thr Lys Lys Asp Ser Gln Thr 130 135 140			
90	Ile Gly Asp Leu Leu Glu Lys Phe Ala Glu Glu Tyr Phe Lys Thr His 145 150 155 160			

48

Lys Arg Thr Thr Lys Ser Glu His Thr Phe Phe Tyr Tyr Phe Ser Arg
 165 170 175
 5 Thr Gln Arg Tyr Thr Asn Ser Lys Asp Leu Ala Thr Ala Glu Asn Leu
 180 185 190
 Ile Asn Ser Ile Glu Gln Ile Asp Lys Glu Trp Ala Arg Tyr Asn Ala
 195 200 205
 10 Ala Arg Ala Ile Ser Ala Phe Cys Ile Thr Phe Asn Ile Glu Ile Asp
 210 215 220
 Leu Ser Gln Tyr Ser Lys Met Pro Asp Arg Asn Ser Arg Asn Ile Pro
 225 230 235 240
 15 Thr Asp Ala Glu Ile Leu Ser Gly Ile Thr Lys Phe Glu Asp Tyr Leu
 245 250 255
 20 Val Thr Arg Gly Asn Gln Val Asn Glu Asp Val Lys Asp Ser Trp Gln
 260 265 270
 Leu Trp Arg Trp Thr Tyr Gly Met Leu Ala Val Phe Gly Leu Arg Pro
 275 280 285
 25 Arg Glu Ile Phe Ile Asn Pro Asn Ile Asp Trp Trp Leu Ser Lys Glu
 290 295 300
 Asn Ile Asp Leu Thr Trp Lys Val Asp Lys Glu Cys Lys Thr Gly Glu
 305 310 315 320
 30 Arg Gln Ala Leu Pro Leu His Lys Glu Trp Ile Asp Glu Phe Asp Leu
 325 330 335
 35 Arg Asn Pro Lys Tyr Leu Glu Met Leu Ala Thr Ala Ile Ser Lys Lys
 340 345 350
 Asp Lys Thr Asn His Ala Glu Ile Thr Ala Leu Thr Gln Arg Ile Ser
 355 360 365
 40 Trp Trp Phe Arg Lys Val Glu Leu Asp Phe Lys Pro Tyr Asp Leu Arg
 370 375 380
 His Ala Trp Ala Ile Arg Ala His Ile Leu Gly Ile Pro Ile Lys Ala
 385 390 395 400
 45 Ala Ala Asp Asn Leu Gly His Ser Met Gln Val His Thr Gln Thr Tyr
 405 410 415
 50 Gln Arg Trp Phe Ser Leu Asp Met Arg Lys Leu Ala Ile Asn Gln Ala
 420 425 430
 Leu Thr Lys Arg Asn Glu Phe Glu Val Ile Arg Glu Glu Asn Ala Lys
 435 440 445
 55 Leu Gln Ile Glu Asn Gln Arg Leu Arg Met Glu Ile Glu Lys Leu Lys
 450 455 460
 Met Glu Ile Ala Tyr Lys Asn Ser
 465 470
 60
 65 <210> 64
 <211> 1008
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> CDS
 <222> (1)..(1005)

5 <220>
 <223> Description of Artificial Sequence: vector
 pBS-SSV3

10 <400> 64
 atg acg aaa gat aag acg cgt tat aaa tac ggg gat tat att tta cgc 48
 Met Thr Lys Asp Lys Thr Arg Tyr Lys Tyr Gly Asp Tyr Ile Leu Arg
 1 5 10 15

15 gag agg aaa ggg cgg tat tat gtt tac aag cta gag tat gaa aac ggt 96
 Glu Arg Lys Gly Arg Tyr Tyr Val Tyr Lys Leu Glu Tyr Glu Asn Gly
 20 25 30

20 gag gta aaa gag cgt tac gtg ggt cct tta got gac gtc gtt gaa tca 144
 Glu Val Lys Glu Arg Tyr Val Gly Pro Leu Ala Asp Val Val Glu Ser
 35 40 45

25 tat cta aaa atg aaa tta ggg gtc gta ggg gat act ccc cta caa gcg 192
 Tyr Leu Lys Met Lys Leu Gly Val Val Gly Asp Thr Pro Leu Gln Ala
 50 55 60

30 gat ccc ccc ggt ttc gag ccc ggg aca agc gga agc ggt ggt gga aaa 240
 Asp Pro Pro Gly Phe Glu Pro Gly Thr Ser Gly Ser Gly Gly Lys
 65 70 75 80

35 gag gga act gaa cga cgt aaa ata gcg ttg gtt gcc aat ttg cgc caa 288
 Glu Gly Thr Glu Arg Arg Lys Ile Ala Leu Val Ala Asn Leu Arg Gln
 85 90 95

40 tac gcg acg gac ggc aac ata aag gcg ttc tac aac tat ctc atg aac 336
 Tyr Ala Thr Asp Gly Asn Ile Lys Ala Phe Tyr Asn Tyr Leu Met Asn
 100 105 110

45 gaa agg ggg ata agc gaa aaa act gca aag gac tac atc aat gct ata 384
 Glu Arg Gly Ile Ser Glu Lys Thr Ala Lys Asp Tyr Ile Asn Ala Ile
 115 120 125

50 tca aag ccg tat aaa gag acg aga gac gca cag aag gct tac cga ctc 432
 Ser Lys Pro Tyr Lys Glu Thr Arg Asp Ala Gln Lys Ala Tyr Arg Leu
 130 135 140

55 ttt gca cgt ttc tta gcg tca cgc aat atc ata cat gat gaa ttt gcg 480
 Phe Ala Arg Phe Leu Ala Ser Arg Asn Ile Ile His Asp Glu Phe Ala
 145 150 155 160

60 gat aaa ata ttg aaa gcg gta aag gtg aag aag gcg aac gct gat atc 528
 Asp Lys Ile Leu Lys Ala Val Lys Val Lys Lys Ala Asn Ala Asp Ile
 165 170 175

65 tac att cca acg ttg gaa gag ata aaa agg acg tta caa tta gca aaa 576
 Tyr Ile Pro Thr Leu Glu Glu Ile Lys Arg Thr Leu Gln Leu Ala Lys
 180 185 190

70 gac tat agc gaa aac gtc tac ttc atc tac cgt atc gct ctc gag tcg 624
 Asp Tyr Ser Glu Asn Val Tyr Phe Ile Tyr Arg Ile Ala Leu Glu Ser
 195 200 205

75 ggc gtt agg ctg agc gaa ata ctg aaa gtg ctg aag gaa ccc gaa agg 672
 Gly Val Arg Leu Ser Glu Ile Leu Lys Val Leu Lys Glu Pro Glu Arg
 210 215 220

80 gac att tgc ggt aac gac gtc tgt tat tat ccg ctt agt tgg act agg 720
 Asp Ile Cys Gly Asn Asp Val Cys Tyr Pro Leu Ser Trp Thr Arg

		50		
	225	230	235	240
5	gga tat aag ggc gtc ttc tat gta ttc cac ata acg cct ctg aag aga Gly Tyr Lys Gly Val Phe Tyr Val Phe His Ile Thr Pro Leu Lys Arg 245 250 255			768
10	gta gag gtg acg aag tgg gca ata gcg gac ttt gaa cga cgt cat aag Val Glu Val Thr Lys Trp Ala Ile Ala Asp Phe Glu Arg Arg His Lys 260 265 270			816
15	gac gct ata gcg ata aag tac ttc cgc aaa ttc gta gcg tct aag atg Asp Ala Ile Ala Ile Lys Tyr Phe Arg Lys Phe Val Ala Ser Lys Met 275 280 285			864
20	gct gag cta agc gta ccg tta gat att atc gat ttt att caa ggg cgt Ala Glu Leu Ser Val Pro Leu Asp Ile Ile Asp Phe Ile Gln Gly Arg 290 295 300			912
25	aaa ccg aca cgc gtt tta acg caa cat tac gta tcg ctc ttc ggc ata Lys Pro Thr Arg Val Leu Thr Gln His Tyr Val Ser Leu Phe Gly Ile 305 310 315 320			960
30	gcg aaa gag caa tat aaa aag tat gcg gaa tgg cta aaa ggg gtc tga Ala Lys Glu Gln Tyr Lys Lys Tyr Ala Glu Trp Leu Lys Gly Val 325 330 335			1008
35	<210> 65 <211> 335 <212> PRT <213> Artificial Sequence <223> Description of Artificial Sequence: vector pBS-SSV3			
40	<400> 65 Met Thr Lys Asp Lys Thr Arg Tyr Lys Tyr Gly Asp Tyr Ile Leu Arg 1 5 10 15			
45	Glu Arg Lys Gly Arg Tyr Tyr Val Tyr Lys Leu Glu Tyr Glu Asn Gly 20 25 30			
50	Glu Val Lys Glu Arg Tyr Val Gly Pro Leu Ala Asp Val Val Glu Ser 35 40 45			
55	Tyr Leu Lys Met Lys Leu Gly Val Val Gly Asp Thr Pro Leu Gln Ala 50 55 60			
60	Asp Pro Pro Gly Phe Glu Pro Gly Thr Ser Gly Ser Gly Gly Lys 65 70 75 80			
65	Glu Gly Thr Glu Arg Arg Lys Ile Ala Leu Val Ala Asn Leu Arg Gln 85 90 95			
70	Tyr Ala Thr Asp Gly Asn Ile Lys Ala Phe Tyr Asn Tyr Leu Met Asn 100 105 110			
75	Glu Arg Gly Ile Ser Glu Lys Thr Ala Lys Asp Tyr Ile Asn Ala Ile 115 120 125			
80	Ser Lys Pro Tyr Lys Glu Thr Arg Asp Ala Gln Lys Ala Tyr Arg Leu 130 135 140			
85	Phe Ala Arg Phe Leu Ala Ser Arg Asn Ile Ile His Asp Glu Phe Ala 145 150 155 160			
90	Asp Lys Ile Leu Lys Ala Val Lys Val Lys Lys Ala Asn Ala Asp Ile 165 170 175			

Tyr Ile Pro Thr Leu Glu Glu Ile Lys Arg Thr Leu Gln Leu Ala Lys
 180 185 190

5 Asp Tyr Ser Glu Asn Val Tyr Phe Ile Tyr Arg Ile Ala Leu Glu Ser
 195 200 205

Gly Val Arg Leu Ser Glu Ile Leu Lys Val Leu Lys Glu Pro Glu Arg
 210 215 220

10 Asp Ile Cys Gly Asn Asp Val Cys Tyr Tyr Pro Leu Ser Trp Thr Arg
 225 230 235 240

Gly Tyr Lys Gly Val Phe Tyr Val Phe His Ile Thr Pro Leu Lys Arg
 15 245 250 255

Val Glu Val Thr Lys Trp Ala Ile Ala Asp Phe Glu Arg Arg His Lys
 260 265 270

20 Asp Ala Ile Ala Ile Lys Tyr Phe Arg Lys Phe Val Ala Ser Lys Met
 275 280 285

Ala Glu Leu Ser Val Pro Leu Asp Ile Ile Asp Phe Ile Gln Gly Arg
 290 295 300

25 Lys Pro Thr Arg Val Leu Thr Gln His Tyr Val Ser Leu Phe Gly Ile
 305 310 315 320

Ala Lys Glu Gln Tyr Lys Tyr Ala Glu Trp Leu Lys Gly Val
 30 325 330 335

35 <210> 66
 <211> 1441
 <212> DNA
 <213> Artificial Sequence

40 <220>
 <223> Description of Artificial Sequence: DNA sequence
 coding for fusion protein NLS-XisA

45 <220>
 <221> CDS
 <222> (1)...(1437)

<400> 66

atg ccc aag aag aag agg aag gtg caa aat cag ggt caa gac aaa tat	48
Met Pro Lys Lys Arg Lys Val Gln Asn Gln Gly Gln Asp Lys Tyr	
1 5 10 15	

50 caa caa gcc ttt gca gac tta gag cca ctt tca tct acc gac ggc agt
 Gln Gln Ala Phe Ala Asp Leu Glu Pro Leu Ser Ser Thr Asp Gly Ser
 20 25 30 96

55 ttt ctc ggc tca agt ctg caa gca cag cag caa aga gaa cac atg aga
 Phe Leu Gly Ser Ser Leu Gln Ala Gln Gln Arg Glu His Met Arg
 35 40 45 144

60 aca aaa gta cta caa gac cta gac aag gta aat ctg cgt ttg aag tct
 Thr Lys Val Leu Gln Asp Leu Asp Lys Val Asn Leu Arg Leu Lys Ser
 50 55 60 192

65 gca aag acg aaa gtc tca gtt cga gaa tct aac gga agt ctg caa tta
 Ala Lys Thr Lys Val Ser Val Arg Glu Ser Asn Gly Ser Leu Gln Leu
 65 70 75 80 240

52

5	cga gca acg tta cca att aaa cct gga gat aag gac acc aac ggt aca Arg Ala Thr Leu Pro Ile Lys Pro Gly Asp Lys Asp Thr Asn Gly Thr 85	90	95	288
10	gac aga aag caa tac aat ctc agc ttg aat atc cct gca aac ttg gat Gly Arg Lys Gln Tyr Asn Leu Ser Leu Asn Ile Pro Ala Asn Leu Asp 100	105	110	336
15	gga ctg aag acg gct gag gaa gct tat gaa tta ggt aaa tta atc Gly Leu Lys Thr Ala Glu Glu Ala Tyr Glu Leu Gly Lys Leu Ile 115	120	125	384
20	gct cgg aaa acc ttt gaa tgg aat gat aaa tat tta ggc aaa gaa gcc Ala Arg Lys Thr Phe Glu Trp Asn Asp Lys Tyr Leu Gly Lys Glu Ala 130	135	140	432
25	act aaa aaa gat tca caa aca ata ggt gat tta cta gaa aaa ttt gca Thr Lys Lys Asp Ser Gln Thr Ile Gly Asp Leu Leu Glu Lys Phe Ala 145	150	155	480
30	gaa gag tat ttt aaa acc cat aaa cgc acc act aaa agc gaa cat acc Glu Glu Tyr Phe Lys Thr His Lys Arg Thr Thr Lys Ser Glu His Thr 165	170	175	528
35	ttt ttt tac tat ttt tcc cgc acc caa cga tat acc aat tcc aaa gat Phe Phe Tyr Tyr Phe Ser Arg Thr Gln Arg Tyr Thr Asn Ser Lys Asp 180	185	190	576
40	tta gca acg gcg gaa aat ctc atc aat tca att gag caa atc gat aaa Leu Ala Thr Ala Glu Asn Leu Ile Asn Ser Ile Glu Gln Ile Asp Lys 195	200	205	624
45	gaa tgg gcg aga tat aat gcc gcc aga gcc ata tca gct ttt tgc ata Glu Trp Ala Arg Tyr Asn Ala Ala Arg Ala Ile Ser Ala Phe Cys Ile 210	215	220	672
50	aca ttc aat ata gaa att gat ttg tcc cag tat tcc aaa atg cct gat Thr Phe Asn Ile Glu Ile Asp Leu Ser Gln Tyr Ser Lys Met Pro Asp 225	230	235	720
55	cgc aat tcg cgc aac atc ccc aca gat gca gaa ata cta tca gga att Arg Asn Ser Arg Asn Ile Pro Thr Asp Ala Glu Ile Leu Ser Gly Ile 245	250	255	768
60	acc aaa ttt gaa gac tat cta gtt acc aga gga aat caa gtt aat gaa Thr Lys Phe Glu Asp Tyr Leu Val Thr Arg Gly Asn Gln Val Asn Glu 260	265	270	816
65	gat gta aaa gat agc tgg caa ctt tgg cgc tgg aca tat gga atg tta Asp Val Lys Asp Ser Trp Gln Leu Trp Arg Trp Thr Tyr Gly Met Leu 275	280	285	864
70	gca gtt ttt ggt tta cgc ccc agg gaa att ttt att aac cct aat att Ala Val Phe Gly Leu Arg Pro Arg Glu Ile Phe Ile Asn Pro Asn Ile 290	295	300	912
75	gat tgg tgg tta agc aaa gag aat ata gac ctc aca tgg aaa gta gac Asp Trp Trp Leu Ser Lys Glu Asn Ile Asp Leu Thr Trp Lys Val Asp 305	310	315	960
80	aaa gaa tgt aaa act ggt gaa aga caa gca tta ccc tta cat aaa gaa Lys Glu Cys Lys Thr Gly Glu Arg Gln Ala Leu Pro Leu His Lys Glu 325	330	335	1008
85	tgg att gat gag ttt gat tta aga aat ccg aaa tat tta gaa atg ctg Trp Ile Asp Glu Phe Asp Leu Arg Asn Pro Lys Tyr Leu Glu Met Leu 340	345	350	1056

gca aca gca att agt aaa aaa gat aaa aca aat cat gct gaa ata aca 1104
 Ala Thr Ala Ile Ser Lys Lys Asp Lys Thr Asn His Ala Glu Ile Thr
 355 360 365

5 gcc tta act cag cgt att agt tgg tgg ttt cgg aaa gtc gaa tta gat 1152
 Ala Leu Thr Gln Arg Ile Ser Trp Trp Phe Arg Lys Val Glu Leu Asp
 370 375 380

10 ttt aaa ccc tat gat tta cgt cac gcc tgg gca atc aga gcg cat att 1200
 Phe Lys Pro Tyr Asp Leu Arg His Ala Trp Ala Ile Arg Ala His Ile
 385 390 395 400

15 tta ggc ata cca atc aaa gcg gcg gct gat aat ttg ggg cat agt atg 1248
 Leu Gly Ile Pro Ile Lys Ala Ala Asp Asn Leu Gly His Ser Met
 405 410 415

20 cag gtt cat aca caa acc tat cag cgc tgg ttc tcg cta gat atg cgg 1296
 Gln Val His Thr Gln Thr Tyr Gln Arg Trp Phe Ser Leu Asp Met Arg
 420 425 430

aag tta gcg att aat cag gct ttg act aag agg aat gaa ttt gag gtg 1344
 Lys Leu Ala Ile Asn Gln Ala Leu Thr Lys Arg Asn Glu Phe Glu Val
 435 440 445

25 att agg gag gag aat gct aaa ttg cag ata gaa aat gaa agg ttg agg 1392
 Ile Arg Glu Glu Asn Ala Lys Leu Gln Ile Glu Asn Glu Arg Leu Arg
 450 455 460

30 atg gaa att gag aag tta aag atg gaa ata gct tat aag aat agt tgag 1441
 Met Glu Ile Glu Lys Leu Lys Met Glu Ile Ala Tyr Lys Asn Ser
 465 470 475

35 <210> 67
 <211> 479
 <212> PRT
 <213> Artificial Sequence
 40 <223> Description of Artificial Sequence: DNA sequence
 coding for fusion protein NLS-XisA

<400> 67
 Met Pro Lys Lys Lys Arg Lys Val Gln Asn Gln Gln Asp Lys Tyr
 1 5 10 15

45 Gln Gln Ala Phe Ala Asp Leu Glu Pro Leu Ser Ser Thr Asp Gly Ser
 20 25 30

50 Phe Leu Gly Ser Ser Leu Gln Ala Gln Gln Arg Glu His Met Arg
 35 40 45

Thr Lys Val Leu Gln Asp Leu Asp Lys Val Asn Leu Arg Leu Lys Ser
 50 55 60

55 Ala Lys Thr Lys Val Ser Val Arg Glu Ser Asn Gly Ser Leu Gln Leu
 65 70 75 80

60 Arg Ala Thr Leu Pro Ile Lys Pro Gly Asp Lys Asp Thr Asn Gly Thr
 85 90 95

Gly Arg Lys Gln Tyr Asn Leu Ser Leu Asn Ile Pro Ala Asn Leu Asp
 100 105 110

65 Gly Leu Lys Thr Ala Glu Glu Ala Tyr Glu Leu Gly Lys Leu Ile
 115 120 125

Ala Arg Lys Thr Phe Glu Trp Asn Asp Lys Tyr Leu Gly Lys Glu Ala

54

130	135	140
5	Thr Lys Lys Asp Ser Gln Thr Ile Gly Asp Leu Leu Glu Lys Phe Ala	
145	150	155
Glu Glu Tyr Phe Lys Thr His Lys Arg Thr Thr Lys Ser Glu His Thr		
165	170	175
10	Phe Phe Tyr Tyr Phe Ser Arg Thr Gln Arg Tyr Thr Asn Ser Lys Asp	
180	185	190
Leu Ala Thr Ala Glu Asn Leu Ile Asn Ser Ile Glu Gln Ile Asp Lys		
195	200	205
15	Glu Trp Ala Arg Tyr Asn Ala Ala Arg Ala Ile Ser Ala Phe Cys Ile	
210	215	220
20	Thr Phe Asn Ile Glu Ile Asp Leu Ser Gln Tyr Ser Lys Met Pro Asp	
225	230	235
Arg Asn Ser Arg Asn Ile Pro Thr Asp Ala Glu Ile Leu Ser Gly Ile		
245	250	255
25	Thr Lys Phe Glu Asp Tyr Leu Val Thr Arg Gly Asn Gln Val Asn Glu	
260	265	270
Asp Val Lys Asp Ser Trp Gln Leu Trp Arg Trp Thr Tyr Gly Met Leu		
275	280	285
30	Ala Val Phe Gly Leu Arg Pro Arg Glu Ile Phe Ile Asn Pro Asn Ile	
290	295	300
35	Asp Trp Trp Leu Ser Lys Glu Asn Ile Asp Leu Thr Trp Lys Val Asp	
305	310	315
Lys Glu Cys Lys Thr Gly Glu Arg Gln Ala Leu Pro Leu His Lys Glu		
325	330	335
40	Trp Ile Asp Glu Phe Asp Leu Arg Asn Pro Lys Tyr Leu Glu Met Leu	
340	345	350
Ala Thr Ala Ile Ser Lys Lys Asp Lys Thr Asn His Ala Glu Ile Thr		
355	360	365
45	Ala Leu Thr Gln Arg Ile Ser Trp Trp Phe Arg Lys Val Glu Leu Asp	
370	375	380
50	Phe Lys Pro Tyr Asp Leu Arg His Ala Trp Ala Ile Arg Ala His Ile	
385	390	395
Leu Gly Ile Pro Ile Lys Ala Ala Asp Asn Leu Gly His Ser Met		
405	410	415
55	Gln Val His Thr Gln Thr Tyr Gln Arg Trp Phe Ser Leu Asp Met Arg	
420	425	430
Lys Leu Ala Ile Asn Gln Ala Leu Thr Lys Arg Asn Glu Phe Glu Val		
435	440	445
60	Ile Arg Glu Glu Asn Ala Lys Leu Gln Ile Glu Asn Glu Arg Leu Arg	
450	455	460
65	Met Glu Ile Glu Lys Leu Lys Met Glu Ile Ala Tyr Lys Asn Ser	
465	470	475

<210> 68

<211> 1029

<212> DNA

<213> Artificial Sequence

5

<220>

<223> Description of Artificial Sequence: DNA sequence
coding for fusion protein NLS-Ssv

10 <220>

<221> CDS

<222> (1)..(1026)

<400> 68

15 atg ccc aag aag aag agg aag gtg acg aaa gat aag acg cgt tat aaa 48
Met Pro Lys Lys Lys Arg Lys Val Thr Lys Asp Lys Thr Arg Tyr Lys
1 5 10 1520 tac ggg gat tat att tta cgc gag agg aaa ggg cgg tat tat gtt tac 96
Tyr Gly Asp Tyr Ile Leu Arg Glu Arg Lys Gly Arg Tyr Tyr Val Tyr
20 25 3025 aag cta gag tat gaa aac ggt gag gta aaa gag cgt tac gtg ggt cct 144
Lys Leu Glu Tyr Glu Asn Gly Glu Val Lys Glu Arg Tyr Val Gly Pro
35 40 4530 tta gct gac gtc gtt gaa tca tat cta aaa atg aaa tta ggg gtc gta 192
Leu Ala Asp Val Val Glu Ser Tyr Leu Lys Met Lys Leu Gly Val Val
50 55 6035 ggg gat act ccc cta caa gcg gat ccc ccc ggt ttc gag ccc ggg aca 240
Gly Asp Thr Pro Leu Gln Ala Asp Pro Pro Gly Phe Glu Pro Gly Thr
65 70 75 8040 agc gga agc ggt ggt gga aaa gag gga act gaa cga cgt aaa ata gcg 288
Ser Gly Ser Gly Gly Lys Glu Gly Thr Glu Arg Arg Lys Ile Ala
85 90 9545 ttg gtt gcc aat ttg cgc caa tac gcg acg gac ggc aac ata aag gcg 336
Leu Val Ala Asn Leu Arg Gln Tyr Ala Thr Asp Gly Asn Ile Lys Ala
100 105 11050 ttc tac aac tat ctc atg aac gaa agg ggg ata agc gaa aaa act gca 384
Phe Tyr Asn Tyr Leu Met Asn Glu Arg Gly Ile Ser Glu Lys Thr Ala
115 120 12555 aag gac tac atc aat gct ata tca aag ccg tat aaa gag acg aga gac 432
Lys Asp Tyr Ile Asn Ala Ile Ser Lys Pro Tyr Lys Glu Thr Arg Asp
130 135 14060 gca cag aag gct tac cga ctc ttt gca cgt ttc tta gcg tca cgc aat 480
Ala Gln Lys Ala Tyr Arg Leu Phe Ala Arg Phe Leu Ala Ser Arg Asn
145 150 155 16065 atc ata cat gat gaa ttt gcg gat aaa ata ttg aaa gcg gta aag gtg 528
Ile Ile His Asp Glu Phe Ala Asp Lys Ile Leu Lys Ala Val Lys Val
165 170 17570 aag aag gcg aac gct gat atc tac att cca acg ttg gaa gag ata aaa 576
Lys Lys Ala Asn Ala Asp Ile Tyr Ile Pro Thr Leu Glu Glu Ile Lys
180 185 19075 agg acg tta caa tta gca aaa gac tat agc gaa aac gtc tac ttc atc 624
Arg Thr Leu Gln Leu Ala Lys Asp Tyr Ser Glu Asn Val Tyr Phe Ile
195 200 205

80 tac cgt atc gct ctc gag tcg ggc gtt agg ctg agc gaa ata ctg aaa 672

56

Tyr Arg Ile Ala Leu Glu Ser Gly Val Arg Leu Ser Glu Ile Leu Lys
 210 215 220

5 gtg ctg aag gaa ccc gaa agg gac att tgc ggt aac gac gtc tgt tat 720
 Val Leu Lys Glu Pro Glu Arg Asp Ile Cys Gly Asn Asp Val Cys Tyr
 225 230 235 240

10 tat ccg ctt agt tgg act agg gga tat aag ggc gtc ttc tat gta ttc 768
 Tyr Pro Leu Ser Trp Thr Arg Gly Tyr Lys Gly Val Phe Tyr Val Phe
 245 250 255

15 cac ata acg cct ctg aag aga gta gag gtg acg aag tgg gca ata gcg 816
 His Ile Thr Pro Leu Lys Arg Val Glu Val Thr Lys Trp Ala Ile Ala
 260 265 270

gac ttt gaa cga cgt cat aag gac gct ata gcg ata aag tac ttc cgc 864
 Asp Phe Glu Arg Arg His Lys Asp Ala Ile Ala Ile Lys Tyr Phe Arg
 275 280 285

20 aaa ttc gta gcg tct aag atg gct gag cta agc gta ccg tta gat att 912
 Lys Phe Val Ala Ser Lys Met Ala Glu Leu Ser Val Pro Leu Asp Ile
 290 295 300

25 atc gat ttt att caa ggg cgt aaa ccg aca cgc gtt tta acg caa cat 960
 Ile Asp Phe Ile Gln Gly Arg Lys Pro Thr Arg Val Leu Thr Gln His
 305 310 315 320

30 tac gta tcg ctc ttc ggc ata gcg aaa gag caa tat aaa aag tat gcg 1008
 Tyr Val Ser Leu Phe Gly Ile Ala Lys Glu Gln Tyr Lys Tyr Ala
 325 330 335

gaa tgg cta aaa ggg gtc tga 1029
 Glu Trp Leu Lys Gly Val
 340

35

<210> 69
 <211> 342
 <212> PRT

40 <213> Artificial Sequence
 <223> Description of Artificial Sequence: DNA sequence
 coding for fusion protein NLS-Ssv

45 <400> 69
 Met Pro Lys Lys Lys Arg Lys Val Thr Lys Asp Lys Thr Arg Tyr Lys
 1 5 10 15

50 Tyr Gly Asp Tyr Ile Leu Arg Glu Arg Lys Gly Arg Tyr Tyr Val Tyr
 20 25 30
 Lys Leu Glu Tyr Glu Asn Gly Glu Val Lys Glu Arg Tyr Val Gly Pro
 35 40 45

55 Leu Ala Asp Val Val Glu Ser Tyr Leu Lys Met Lys Leu Gly Val Val
 50 55 60

60 Gly Asp Thr Pro Leu Gln Ala Asp Pro Pro Gly Phe Glu Pro Gly Thr
 65 70 75 80

65 Ser Gly Ser Gly Gly Lys Glu Gly Thr Glu Arg Arg Lys Ile Ala
 85 90 95

Leu Val Ala Asn Leu Arg Gln Tyr Ala Thr Asp Gly Asn Ile Lys Ala
 100 105 110

Phe Tyr Asn Tyr Leu Met Asn Glu Arg Gly Ile Ser Glu Lys Thr Ala
 115 120 125

Lys Asp Tyr Ile Asn Ala Ile Ser Lys Pro Tyr Lys Glu Thr Arg Asp
 130 135 140
 5 Ala Gln Lys Ala Tyr Arg Leu Phe Ala Arg Phe Leu Ala Ser Arg Asn
 145 150 155 160
 Ile Ile His Asp Glu Phe Ala Asp Lys Ile Leu Lys Ala Val Lys Val
 10 165 170 175
 Lys Lys Ala Asn Ala Asp Ile Tyr Ile Pro Thr Leu Glu Glu Ile Lys
 180 185 190
 15 Arg Thr Leu Gln Leu Ala Lys Asp Tyr Ser Glu Asn Val Tyr Phe Ile
 195 200 205
 Tyr Arg Ile Ala Leu Glu Ser Gly Val Arg Leu Ser Glu Ile Leu Lys
 210 215 220
 20 Val Leu Lys Glu Pro Glu Arg Asp Ile Cys Gly Asn Asp Val Cys Tyr
 225 230 235 240
 Tyr Pro Leu Ser Trp Thr Arg Gly Tyr Lys Gly Val Phe Tyr Val Phe
 245 250 255
 25 His Ile Thr Pro Leu Lys Arg Val Glu Val Thr Lys Trp Ala Ile Ala
 260 265 270
 Asp Phe Glu Arg Arg His Lys Asp Ala Ile Ala Ile Lys Tyr Phe Arg
 30 275 280 285
 Lys Phe Val Ala Ser Lys Met Ala Glu Leu Ser Val Pro Leu Asp Ile
 290 295 300
 35 Ile Asp Phe Ile Gln Gly Arg Lys Pro Thr Arg Val Leu Thr Gln His
 305 310 315 320
 Tyr Val Ser Leu Phe Gly Ile Ala Lys Glu Gln Tyr Lys Lys Tyr Ala
 325 330 335
 40 Glu Trp Leu Lys Gly Val
 340

45 <210> 70
 <211> 3908
 <212> DNA
 50 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: vector
 pBS-SSV3

55 <400> 70
 cacctaaatt gtaagcgtta atattttgtt aaaattcgcg ttaaaatttt gttttatcag 60
 ctcattttt aaccaatagg ccggaaatcg caaaatccct tataaatcaa aagaatagac 120
 cgagataggg ttgagtgttgc ttccagtttgc gaacaagagt ccactattaa agaacgtgg 180
 ctccaaacgttc aaaggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc 240
 60 accctaatac agttttttgg ggtcgagggtt ccgtaaagca ctaaatcgga accctaaagg 300
 gagcccccgaa tttagagctt gacggggaaa gccggcgaaac gtggcgagaa aggaaggaa 360
 gaaagcgaa ggagcgccgctt cttagggcgctt ggcaagtgttgc ggggtcacgc tgccgttaac 420
 caccacaccc gcccgcgtta atgcgcgcgtt acaggcgccg tcccatcgcc cattcagggt 480
 65 gcgcaactgtt tggttggaaa gatcggtcg ggccttctcg ctattacgccc agctggcgaa 540
 agggggatgtt gctgcaaggc gattaatggt ggttacgc gggttttccc agtcacgc 600
 ttgtaaaacg acggccagtg aattgtataa cgactcaacta tagggcgaaat tgtagctcca 660
 ccgcgggtggc ggccgcccga tatgacgaaa gataagacgc gttataaataa cggggattat 720

attttacgcg agaggaaagg gcggtattat gttacaagc tagagtatga aaacggtag 780
 gtaaaagagc gttacgtggg tccttagct gacgtcggt aatcatatct aaaaatgaaa 840
 ttaggggtcg tagggatac tcccctacaa gcggtatccc ccggttcga gcccgggaca 900
 5 agcggaaagcg gtggggaaa agagggact gaacgacgta aaatagcggt gttgccaat 960
 ttgcgccaa atcgacgga cggcaacata aaggcggt acaactatct catgaacgaa 1020
 agggggataa gcgaaaaaac tgcaaaggac tacatcaatg ctatataaa gccgtataaa 1080
 gagacgagag acgcacagaa ggcttaccga ctcttgcac gtttcttagc gtcacgcaat 1140
 atcatacatg atgaatttgc ggataaaaata ttgaaagcgg taaaggtaga gaaggcgaac 1200
 gctgatatct acatccaaac ttggaagag ataaaaagga cgttacaatt agcaaaagac 1260
 10 tatagcgaaa acgtctactt catctaccgt atcgctctcg agtcggcgt taggctgagc 1320
 gaaatactga aagtgtgaa ggaaccggaa agggacattt gcggtaaacga cgtctgttat 1380
 tatccgctt gttgacttag gggatataag gggtcttct atgtattcca cataacgcct 1440
 ctgaagagag tagaggtgac gaagtggcga atagcgact ttgaacgacg tcataaggac 1500
 gctatacgca taaagtactt ccgcaattt gtagegtcta agatggctga gctaagcgt 1560
 15 ccgttagata ttatcgattt tacgtatcgc tcttcggcat tattcaaggc caccgcgttt aacgcaacat 1620
 ggggtctgac tcgagggggg gggatataaa agtatgcgg aatggctaaaa 1680
 ttcgagctt gcgtaatcat ggtcatagct gtttcctgtg taaaattgtt atccgctcac 1800
 aattccacac aacatacgag ccggaaagcat aaagtgtaaa gcctggggc ctaatgagt 1860
 20 gagctaactc acattaattt gtcgcagctg cattaatgaa tcggccaaacg cgcggggaga ggcgggttgc gtattggcg 1920
 ctcttccgct tcctcgctca ctgactcgct ggcgtcggtc gttcggctgc ggccgagcgt 2040
 atcagctcac tcaaaggcgg taataacggtt atccacagaa tcagggata acgcagggaa 2100
 25 gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cggtgttgc 2160
 gttttccat aggctccgc cccctgacga gcatcacaaa aatcgacgt caagtcagag 2220
 gtggcgaaccc ccgacaggac tataaagata ccaggcggtt cccctggaa gtcctctgt 2280
 gcgctctctt gttccgaccc tgccgttac cggatacctg tccgccttc tcccttcggg 2340
 aagcgtggcg ctttctcata gtcacgctg tagtatctc agttcggtgt aggtcggtc 2400
 30 ctccaagctg ggctgtgtgc taactatctg cttgagtcca tggtaacagg attacgagag
 gcttaactac ggctacaacta tacccctcgaa aaaaggtt ggtttttt gtttgcga
 35 tttgatctt tctacggggt ggtcatgaga ttatcaaaaa taaatcaatc taaagtatat
 tgaggcacct atctcagcga cgttagata actacgatac gcgagacca
 40 gcgagacca cgctcaccgg cgagcgcaga agtgtcccg ggaagctaga gtaagttagt
 aggcatcggt gtgtcacgct atcaaggcga gttacatgtatccatcgatcgtt
 45 tccgatcggt gtcagaagta gttggccgc agtgttatca gcatcgatcgtt
 gcataattct cttactgtca tgccatccgt aagatgtttt tctgtactg gtgagactc
 aaccaagtca ttctgagaat agtgtatcg ggcaccgat
 50 acgggataat accgcggccat ttccggggcga aaactctcaa tcgtgcaccc
 aactgatctt aacaggaagg caaaatgccc catactcttc tttttcaat
 atacatattt gaatgtattt aaaaataaa acaaatacggtt
 55 aaaaatgca 3908

<210> 71
 <211> 3927
 <212> DNA

60 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: vector
 pBS-SSV4

65 <400> 71
 cacctaaattt gtaagcgta atattttgtt aaaattcgcc taaaattttt gttaaatcag 60

5 ctcattttt aaccaatagg ccgaaatcg caaaatccct tataaatcaa aagaatagac 120
 cgagataggg ttgagtgtt ttccagttt gaacaagagt ccactattaa agaacgtgga 180
 ctccaacgtc aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc 240
 accctaatca agtfffftgg ggtcgaggtg cctaaatcgga accctaaagg 300
 gagcccccga titagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa 360
 gaaagcgaaa ggagcggcg ctagggcgct ggcaagtgtt gcggtcacgc tgcgcgttaac 420
 caccacaccc gccgcgtta atgcgcgc acaggcgcg tcccattcgc cattcaggct 480
 gcgcaactgt tgggaaggc gatcggtcg ggccttctcg ctattacgcc agctggcgaa 540
 10 agggggatgt gctgcaaggc gattaagtg ggttaacgca ggttttccc agtcacgacg 600
 ttgtaaaaacg acggccagtg aatitgtataa cgactcaacta tagggcgaat tggagctcca 660
 cgcgggtggc gcccgcacca tgcccaagaa gaagaggaag gtgacgaaag ataagacgcg 720
 ttataaaatc ggggattata ttttacgcga gaggaaaggc cggtattatg tttacaagct 780
 agagtatgaa aacggtgagg taaaagagcg ttacgtgggt cctttagctg acgtcggttg 840
 15 atcatatcta aaaaatgaaat taggggtcg aggggatact cccctacaag cggatcccc 900
 cggtttcgag cccgggacaa gcggaagcg tggggaaaaa ggggaactg aacgacgtaa 960
 aatagcggtg gttggcaatt tgcccaata cgcgacggac ggcaacataa aggcgttcta 1020
 caactatctc atgaacgaaa gggggataaag cggaaaaact gcaaaggact acatcaatgc 1080
 tataatcaaag ccgtataaaag agacgagaga cgcacagaag gcttaccgac tctttgcacg 1140
 20 ttttttagcg tcacgcaata tcatacatga tgaatttgcg gataaaaat taaaagcggt 1200
 aaagggtgaag aaggcgaacg ctgatatact aatttccaaacg ttggaaagaga taaaaggac 1260
 gttacaattn gcaaaaagact atagcgaaaaa cgtctacttc atctaccgtt tgcgtctcg 1320
 gtcggggcggtt aggctgagcg aaatactgaa agtgcgtgaag gaacccgaaa gggacattt 1380
 cggtaacgac gtctgttatt atccgcttag ttggactagg ggtatataagg gcgctttcta 1440
 25 tgtattccac ataacgcctc tgaagagagt agagggtacg aagtggcaaa tagcggactt 1500
 tgaacgacgt cataaggacg ctatagcgat aaagtacttc cggaaatttcg tagcgtctaa 1560
 gatggctgag ctaacgcgtac cgttagat tatcgatttt attcaaggcgt gtaaaccgac 1620
 acgcgttttta acgcaacatt acgtatcgct cttcggcata gcaaaagagc aatataaaaa 1680
 gtagcggaaa tggctaaaag gggctctgact cgagggggggg cccggtaccc agcttttgtt 1740
 30 cccttttagt ggggttaatt tcggagcttgg cgtaatcatg gtcatacgctg tttctgtgt 1800
 gaaattttgtt tccgctcaca attccacaca acatcagcgc cggaaagcata aagtgtaaag 1860
 cctgggggtgc ctaatgagtg agctaactca cattaattgc tttccggcgtt cttccgtac tgactcgctg 1920
 tccagtcggg aaacctgtcg tggcagctgc attaatgaat cggcaacgcg gggggagag 1980
 35 gcggtttgcg tattttggcgc ttccggctcg gcgagcggta tcagctcaact caaaggcgtt aatacggtt tccacagaat 2100
 caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 2160
 aaaaggccgc gttgtggcg ttttccata ggctccgccc ccctgacgag catcacaaaa 2220
 atcgacgcgc aagtctcgagg tggcgaaacc cgacaggact ataaagatac caggcggttc 2280
 cccctggaaag ctccctcggt cgctctcctg ttccgaccct gccgcttacc ggataccctgt 2340
 40 ccccttttct cccttcggga agcgtggcgc ttctctatag ctcacgctgt aggtatctca 2400
 gttcgggtgtt ggtcggttgc tccaaagctgg gctgtgtgcg cgaacccccc gttcagccg 2460
 accgcgtgcgc ctatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 2520
 cgccactgtgc agcagccact ggttaacagga ttacgagacg gaggtatgtt ggcgggtgcta 2580
 cagagtctt gaagtgggtt cctaactacg gctacactag aaggacagta ttgtgtatct 2640
 45 ggcgtctgtt gaaggcaggta accttcggaa aaagagttgg tagctctgtt tccggcaaac 2700
 aaaccacccgc tggtagcggt ggttttttgg tttgcaagca gcaagattacg cgcagaaaaa 2760
 aaggatctca agaagatctt actcacgtt agggattttt gtcatgagat tatcaaaaag gatcttcacc tagatccctt 2820
 taaattaaaa atgaatgttt aatcaatct aaatgtatata tgagttaaact ttgtctgaca 2880
 gttaccaatg cttaatcgtt gaggcaccta totcagcgat ctgtcttattt ctttcacca 3000
 50 tagttgcctg actccccgtc gtttagataa ctacgatacg ggagggctta ccatctggcc 3060
 ccagtctgtc aatgataccg cgagaccac gtcacccgc tccagattt tcaacatcaa 3120
 accagccagc cggaaaggccc agtctattaa ttgttgcggg gagcgttagag taatgttgc tttttttttt 3180
 acgttgcgttgc cattgtctaca tcaaggcggag ttacatgatc cccatgttg tgcaaaaaa 3240
 55 tcaaggcggag ttcccaacga cggtagctc ctccggctt tcaaggcggag ttacatgatc cccatgttg tgcaaaaaa 3360
 tcatgggtt ggcagcactg ctgtgactgg tgagtagtca gtcacccgc tttttttttt 3420
 gcttgcggcc ggcgtcaata cttccatgtt gtcacccgc tttttttttt 3480
 60 tcatcatgtt aaaaacgttct cttccatgtt gtcacccgc tttttttttt 3540
 ccagttcgat gtaacccact cttccatgtt gtcacccgc tttttttttt 3600
 gggatataa cttccatgtt gtcacccgc tttttttttt 3660
 tcatcatgtt aaaaacgttct cttccatgtt gtcacccgc tttttttttt 3720
 65 gggatataa cttccatgtt gtcacccgc tttttttttt 3780
 cttccatgtt gtcacccgc tttttttttt 3840
 gttattgtt gtcacccgc tttttttttt 3900
 ttccgcgcac attttttttt 3927

<210> 72
 <211> 3351
 <212> DNA
 <213> Artificial Sequence
 5
 <220>
 <223> Description of Artificial Sequence: vector
 pBS-SSVs
 10 <400> 72
 cgacctcgag tcagaccctt tttagccatt ccgcatactt tttatattgc tcttcgcta 60
 tgccgaagag cgatacgtaa tggtcggtt aaacgcgtgt cggttacgc ccttgaataa 120
 aatcgataat atctaacggt acgcttagct cagccatctt agacgcgtacg aatttgcgga 180
 agtactttt cgctatagcg tcctttagac gtcgttcaaa gtcgttattt gcccacttcg 240
 15 tcacctctac tctttcaga ggcgtttagt ggaatacata gaagacgccc ttatatccc 300
 tagtccaact aagggtataa taacagacgt cgttaccgca aatgtccctt tccgggttc 360
 tcagcactt cagttttcg gttttgttc ccttttagtga gggtaattt cgagttggc gggggggc 420
 ttccctgtgtg aaattgttat ccgttcacaa ttccacacaa gtaatcatgg tcatagctg 480
 20 agtgtaaagc ctgggtgtcc taatgagtga gcttaactcac catacgagcc ggaagcataa 540
 tgcccgttt ccagtcggga aacctgtcg gccaagctgca attaattgc ttgcgtcac 600
 cggggagagg cggttgcgt gctcggtcg tccgtgcgg ccacagaatc agggataac 660
 25 ggaaccgtaa aaaggcccg attcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc 720
 aggcgtttcc cccttggaaagc gataacctgc cgccttctc cttcggtat cagtcactc 780
 gataacctgc cgccttctc ggtatctcg ttcgtgttag gtcgttgcgtt ccaagctggg 840
 30 ttcagcccgaa ccgtcgcc acgacttatac gccactggca gtcgttgcgtt gtcgttgcgtt 900
 gccgtgtcgtc agagttcttg ttcgttgcgtt gtcgttgcgtt gtcgttgcgtt 960
 attcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc 1020
 35 agtcacccatc cccttggaaagc gataacctgc cgccttctc gtcgttgcgtt gtcgttgcgtt 1080
 gataacctgc cgccttctc gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1140
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1200
 40 ttcagcccgaa ccgtcgcc acgacttatac gccactggca gtcgttgcgtt gtcgttgcgtt 1260
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1320
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1380
 45 ttcagcccgaa ccgtcgcc acgacttatac gccactggca gtcgttgcgtt gtcgttgcgtt 1440
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1500
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1560
 50 ttcagcccgaa ccgtcgcc acgacttatac gccactggca gtcgttgcgtt gtcgttgcgtt 1620
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1680
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1740
 55 ttcagcccgaa ccgtcgcc acgacttatac gccactggca gtcgttgcgtt gtcgttgcgtt 1800
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1860
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 1920
 60 ttcagcccgaa ccgtcgcc acgacttatac gccactggca gtcgttgcgtt gtcgttgcgtt 1980
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2040
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2100
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2160
 65 ttcagcccgaa ccgtcgcc acgacttatac gccactggca gtcgttgcgtt gtcgttgcgtt 2220
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2280
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2340
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2400
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2460
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2520
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2580
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2640
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2700
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2760
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2820
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2880
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 2940
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 3000
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 3060
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 3120
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 3180
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 3240
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 3300
 gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 3360

<210> 73

<211> 5730

<212> DNA

5 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pCMVC31(NNLS)

10

<400> 73

aaacagtcg atgtacgggc cagatatacg cgttgcattt gattattgac tagttattaa 60
 tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg ctttacataa 120
 cttacggtaa atggcccggc tggctgaccg cccaacgacc cccgcccatt gacgtcaata 180
 15 atgacgtatg ttccatagg aacgccaata gggactttcc attgacgtca atgggtggac 240
 tatttacgtt aaactgccc cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
 cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgacctt 360
 tgggacttcc tctactggca gtacatctac gtattagtca tcgctattac catggtgatg 420
 20 cgggtttggc agtacatcaa tgggcgttga tagcggtttg actcacgggg atttccaagt 480
 ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca 540
 aaatgtcgta acaactccgc cccattgacg caaatggcg gttaggcgtgt acgggtggag 600
 gtctatataa gcagagctct ctggcttaact agagaaccca ctgcttactg gcttatcgaa 660
 attaatacga ctcaactatag ggagacccaa gctgactcta gacttaatta aeggttgggg 720
 25 ttagtactcc ctctcaaaaag cgaggaggat ttgatattca ctggtcagaa aagacaatct ttttgggttc aagcttgagg tggcggcaggc ttgagatctg 900
 gccatacact tgagtgacat tgacatccac tttgccttcc tctccacagg tggccactcc 960
 cagggcggcc gcacccatgcc caagaagaag aggaaggta cacaagggggt tggaccggg 1020
 30 gtggacacgt acgggggtgc ttacgaccgt cagtcgcgcg agcgcgagaa ttgcagcgca 1080
 gcaaggcccag cgacacacgc taggcggccaa gaagacaagg cggccgcacct tcagcgcgaa 1140
 gtcgagcgcg acggggggcg gttcagggttc gtcgggcatt tcagcgaaagc gccgggcacg 1200
 tcggcgttgc ggacggcggaa 35 gggcggctca acatgatcat tgcgtatgac gtgtcgcgct ttcgcgcctt gaaggtcatg 1320
 gacgcgattc cgattgtctc ggaattgctc ggcctggcg tgacgattgt ttccactcag 1380
 gaaggcgctt tccggcaggg aaacgtcatg gacctgattc acctgattat gcccgtcgac 1440
 gctgcgcaca aagaatctt gctgaagtgc gegaagattc tcgacacacgaa gaaccttca 1500
 cgcgaattgg gccccgtacgt 40 acgaaggaga tcacgcgca tgcaccactc cccttaccgg acacccatccc ttcaagccgg gcagtcgaagc cgcatttcac 1740
 cgtgagatca agacgcacaa cccggcagca tcacgggct ttgtaaagcgc atggacgctg acggcggtcc gaccggggc 1800
 gagacgattt ggaagaagac cgcttcaagc gcctgggacc cggcaaccgt tatgcgaatc 1860
 cttcgggacc cgcgtattgc 45 gggcttcgca gctgagggtac tgcatttcgcg gcccgtccgg 1980
 ggcacgcgcg ccacgaagat cccgtcgac ttgttggcgg tggcaaggggccttct tctgtccgc 2100
 cccgtcgac ttgttggcgg 50 atcaaggact ctiaccgcgt caccatggcg gcaactcgaca agttcggttc ggaacgcata 2280
 cacgaaggca cgtgcacatc ttcaacaaga tcaggcacgc cgaaggccgac 2340
 gcccgcacgt tcggcaagct 55 cactgaggcg cctgagaaga gcggcgaacg ggcgaacctt 2400
 gttgcggagc ggcgcgcacgc cctgaacgc cttgaagagc tgcgtcgcc 2460
 ggcgcgtacg acggaccgcgt 60 tggcagggaaac cacttccgaa agcaacaggc agcgctgacg 2520
 ctccoggcagc aagggggcgg 55 agagcggctt gccgaacttgc aagccgcga agccccggaa 2580
 ctccccctt accaatggtt tgggtggggc ggcgcgtcactt acccgaccgg ccctaagtgc 2640
 tgggtggggc ggcgcgtcactt 65 atcgttgtca cgaagtcgac gacgcacgc gacgcacgcacg 2820
 tcgatcacgt gggcgaagcc ggcgcgtcactt ggcgcgtcactt acccgaccgg ccctaagtgc 2700
 gaagacgttag cggcgttaggc 60 tcaatgtatct ttgtgtttgtt tttttgtgt ctttggggaa gggggaggcc 2940
 tcattatttt caatgtatct ggcgcgtcactt ggcgcgtcactt acccgaccgg ccctaagtgc 3000
 agaatgagggc ggcgcgcacgc 65 cggggggggc ggcgcgtcactt acccgaccgg ccctaagtgc 3060
 cagaatgacc ttggggggggc gggaggccag aatgaggccgc gcccccgggt accgagctcg 3120
 aatttcactgg ccgtcggttt 70 acaacgtcgta gacttggaaa accctggcgta tacccttactt ggcgcgcacc 3180
 aatcgccctt cggccatcc 75 gatcgccctt cccacatcgat ggcgcgtcactt acccgaccgg ccctaagtgc 3240
 ctccttacgc atctgtgcgg 80 tatttcacac cggccatcgat ggcgcgtcactt acccgaccgg ccctaagtgc 3300
 tctgtatgcg catagttaa ccagccccga caccgcacaa caccgcgtga cgcgccttgc 3360

5 cgggcttgtc tgctccggc atccgttac agacaagctg tgaccgtc cgggagctgc 3420
 atgtgtcaga gttttcacc gtcataccg aaacgcgcga gacgaaaggg cctcgtata 3480
 cgcctatttt tataggtaa tgtcatgata ataatggttt cttagacgtc aggtggcaet 3540
 ttcggggaa atgtgcgcgg aacccttatt tgtttatttt tctaaataca ttcaaataatg 3600
 10 tatccgctca tgagacaata accctgataa atgcttaat aatattgaaa aaggaagagt 3660
 atgagttttt aacatttccg tgccgcctt attccctttt ttgcggcatt ttgccttcct 3720
 gttttgttc acccagaaaac gctgggtaaa gtaaaagatg ctgaagatca gtgggggtca 3780
 cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag tttcgcccc 3840
 gaagaacgtt ttccaatgat gaggactttt aaagtttgc tatgtggcgc ggtattatcc 3900
 15 cgtattgacg cccggcaaga gcaactcggt cgccgcatac actattctca gaatgacttg 3960
 gtggagtact caccagtac agaaaagatc cttacggatg gcatgacagt aagagaattt 4020
 tgcagtgtc ccataaccat gaggataaac actgcggcca acttactct gacaacgatc 4080
 ggaggaccga aggagctaac cgctttttt cacaacatgg gggatcatgt aactcgccct 4140
 gatcgttggg aaccggagct gaatgaagcc ataccaaacc acgagcgtga caccacgatg 4200
 20 cctgttagcaa tggcaacaac gttgogcaaa ctatataactg gcgaaactact tactcttagct 4260
 tccggcaac aattaataga ctggatggag gcgatataaag ttgcaggacc acttctgcgc 4320
 tccggccctt cggctggctg gtttattgt gataaaatctg gacccgtgt gctgggggtct 4380
 cgcgttatac ttgcagcact gggccagat ggttaagccct cccgtatgtt agttatctac 4440
 acgacggggg gtcaggcaac tatggatgaa cgaatagac agatcgtga gatagggtgcc 4500
 25 tcaactgatta agcattggta actgtcagac caagttact catatataact tttagattgt 4560
 ttaaaaacttc atttttaatt taaaaggatc tagtgaaaga tcccttttga taatctcatg 4620
 accaaaaatcc cttaacgtga gtttcgttc cactgagcgt cagaccctgtt agaaaagatc 4680
 aaaggatctt cttagagatcc ttttttctg cgcttaatct gctgcttgca aaaaaaaaaa 4740
 30 ccaccgctac caggggttgtt ttggttggcc gatcaagagc taccacactt ttttccgaag 4800
 gtaactggct tcagcagagc gcagatacca aataactgtcc ttcttagtggta gccgttagtta 4860
 gcccaccact tcaagaactc ttagcaccg cctacatacc tgcgtctgt aatcctgtt 4920
 ccagtggctg ctgcccagtgg cgataagtcg tgccttaccg gtttggactc aagacgatag 4980
 ttacccggata aggcgcagcg gtcgggctga acgggggtt cgtgcacaca gcccagcttg 5040
 35 gagcgaacgaa cctacaccga actgagatac ctacagcgtg agctatgaga aagcgcac 5100
 ctcccgaag ggagaaaaggc ggacaggat cccgttaagcg gcaggggtcgg aacaggagag 5160
 cgcacgaggg agcttccagg gggaaacgccc tggtatctt atagtctgtt cgggtttcgc 5220
 caccctgtac ttgagcgtcg attttgcata tgctcgtag gggggccggag cctatggaaa 5280
 aacgcgcagca acgcggccctt ttacgggtt ctgcctttt gctggccctt tgctcacatg 5340
 ttctttctgt cgttatcccc tgattctgt gataaccgtt ttaccggctt tgagtggact 5400
 40 gataccgctc gcccgcagccg aacgacccgg cgcacgcgtt cagtggcgaa ggaagcggaa 5460
 gagcgcggcaaa taacgcacccaa gcctctcccc ggcgttggc cgattcatta atgcagctgg 5520
 cacgacaggt ttcccgactg gaaagcgggc agtgagcgtca acgcaattaa tttatgttgc 5580
 ctcaactcattt aggacacccca ggcttacac tttatgttgc cggctcgat gttgtgtgg 5640
 attgtgagcg gataacaattt tcacacagga aacagctatg accatgatta cgccaaagcta 5700
 gccccggctt gcttgcatttgc ctgcagggtt 5730

45 <210> 74
 <211> 4886
 <212> DNA
 <213> Artificial Sequence

50 <220>
 <223> Description of Artificial Sequence: vector
 pCMV-SSV

55 <400> 74
 aacacagtccg atgtacgggc cagatatacg ctgttgcattt gattattgac tagtttattaa 60
 tagtaatcaa ttacggggtc attagttcat agccatata tggagttccg ctttacataaa 120
 cttacggtaa atggcccgcc tggctgaccg cccaaacgacc cccgcccatt gacgtcaata 180
 atgacgtatg ttcccatatgt aacgccaata gggactttcc attgacgtca atgggtggac 240
 tattttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
 cctatttgcac tcaatgcacgg taaatggccc gcctggcatt atgcccagta catgaccctt 360
 tggacttttcc tcaacttggca gtacatctac gtattttgtca tgcgttattac catggtgatg 420
 60 cgttttggc agtacatcaa tggcgttggc tagcggtttgc actcacgggg atttccaagt 480
 ctccacccca ttgacgtcaa tggagtttgc ttgttggcacc aaaatcaacg ggacittcca 540
 aatgtcgta acaacccgc cccatttgcac caaatggccg gtggcgtgt acgggtggag 600
 gtcataataa gcagagctct ctggctactt agagaacccaa ctgcttactg gcttatacgaa 660
 attaataacga ctcaactatag ggagacccaa gctgactcta gacttaatta agcgttgggg 720
 65 tgagtactcc ctctcaaaaag cggccatgac ttctgcgtca agattgtcag ttccaaaaaa 780
 cgaggaggat ttgatattca cctggcccgcc ggtgatgcct ttgagggtgg cccggtccat 840
 ctggtcagaa aagacaatct ttttgttgc aagcttgagg tggatctgttgc 900

	gccatacact	tgagtgacat	tgacatccac	tttgccttgc	tctccacagg	tgtccactcc	960
	caggcgccc	gcccgatatg	acgaaaagata	agacgcgita	taaatacggg	gattatattt	1020
	tacgcgagag	gaaagggcgg	tattatgtt	acaagctaga	gtatgaaaac	ggtaggtaa	1080
5	aagagcgtta	cgtgggtcct	ttagctgacg	tcgttgaatc	atatctaaaa	atgaaaattag	1140
	gggtcgtagg	ggatactccc	ctacaagcgg	atccccccgg	tttcgagccc	gggacaagcg	1200
	gaagcggtagg	tggaaaagag	ggaactgaac	gacgtaaaat	agcgttggtt	gccaatttgc	1260
	gccaatacgc	gacggacggc	aacataaagg	cgttctacaa	ctatctcatg	aacgaaaagg	1320
	ggataagcga	aaaaactgca	aaggactaca	tcaatgttat	atcaaagccg	tataaagaga	1380
10	cgagagacgc	acagaaggct	taccgactct	ttgcacgttt	cttagcgtca	cgcaatatac	1440
	tacatgatga	atttgcggat	aaaatattga	aagcggtaaa	ggtgaagaag	gccaacgctg	1500
	atatctacat	tccaaacgtt	gaagagataa	aaaggacgtt	acaattagca	aaagactata	1560
	gcgaaaacgt	ctacttcatc	taccgtatcg	ctctcgagtc	gggcgtttagg	ctgagcgaaa	1620
	tactgaaagt	gctgaaggaa	cccgaaaagg	acatttgcgg	taacgcgtc	tgttatttac	1680
15	cgcttagtt	gactagggg	tataaggcg	tcttctatgt	attccacata	acgcctctga	1740
	agagagttaga	ggtgacgaaag	tgggcaatag	cggactttga	acgacgtcat	aaggacgcta	1800
	tagcgataaa	gtacttcggc	aaatctcgtag	cgtctaatag	ggctgagcta	acgcttacccgt	1860
	tagatattat	cgattttat	caagggcgta	aaccgcacg	cgttttaaacg	caacattacg	1920
	tatcgctctt	cggcatacg	aaagagcaat	ataaaaagta	tgcggaatgg	ctaaaagggg	1980
20	tctgactcga	ggggggggcc	gtcgacctcg	agatccaggc	gcccgttacat	aaaagatcat	2040
	tattttcaat	agatctgtgt	gttgggtttt	tgtgtgcctt	gggggggggg	gaggccagaa	2100
	tgaggcgccc	ccaaggggg	ggggggagcc	agaatgacct	tggggggaggg	ggaggccaga	2160
	atgaccttgg	gggaggggg	ggccagaatg	agggcgcccc	ccgggttacccg	agctcgaatt	2220
	caactggccgt	cgttttacaa	cgtcgtgact	ggggaaaaccc	tggcggttacc	caacttaatc	2280
25	gccttgcagc	acatccccct	ttcgcccagct	ggcgtataatag	cgaagaggcc	cgcaccgatc	2340
	gcccctccca	acagttgcgc	agcctgaatg	gcgaatggcg	cctgatgcgg	tattttctcc	2400
	ttaecgcacat	gtgcggatt	tcacaccgca	tatggtgccac	tctcagttaca	atctgctctg	2460
	atgcccgcata	gttaagccag	ccccgacacc	cgccaacacc	cgctgacgcgg	ccctgacccgg	2520
	cttgcctctgt	cccgccatcc	gcttacagac	aagctgtgac	cgctctccgg	agctgcacgt	2580
30	gtcagaggtt	ttcacccgtca	tcaccgaaac	gcccggagacg	aaagggcctc	gtgataccgc	2640
	tatttttat	ggtaatgtc	atgataat	tggtttctta	gacgtcagg	ggcacttttc	2700
	ggggaaatgt	gcccggaaacc	cctattttgtt	tattttttata	aatacattca	aatatgtatc	2760
	cgcctcatgag	acaataaacc	tgataaatgc	ttaataataa	ttggggaaaagg	aagagatgt	2820
	gtattcaaca	tttccgtgtc	gcccatttttc	ccttttttgc	ggcatttttc	cttccctgttt	2880
35	ttgctcaccc	agaaaacgtg	gtgaaaagtaa	aagatgtgta	agatcaatgt	ggcgtacccgg	2940
	tgggttacat	cgaactggat	ctcaacagcg	gtaagatctt	tgagagttt	cgccccgaag	3000
	aacgttttcc	aatgtatgagc	acttttaaag	ttctgctatg	tggcgccgt	ttatcccgta	3060
	ttgacgcggg	gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaaat	gacttggttt	3120
	agtaactcacc	agtcacagaa	aagcatctt	cggtatggcat	gacagtaaga	gaattatgca	3180
40	gtgctccat	aaccatgagt	gataacactg	cggccaactt	acttctgaca	acgatccggag	3240
	gaccgaaggg	gctaaccgct	tttttgcaca	acatggggga	tcatgttaact	cgccttgcac	3300
	gttgggaacc	ggagctgaat	gaagccatac	caaacgcgca	gcgtgacacc	acgatgcctg	3360
	tagcaatggc	aacaacgtt	cgaaaactat	taactggcg	actacttact	ctagcttccc	3420
	ggcaacaatt	aatagactgg	atggaggccg	ataaaagttgc	aggaccactt	ctgcgtcgcg	3480
45	cccttccggc	tggctggttt	attgctgata	aatctggagc	cggtgacgcgt	gggtctcgcg	3540
	gtatcattgc	agcactgggg	ccagatgtta	agccctcccg	tatcgtagtt	atctacacga	3600
	cggggagtc	ggcaactatg	gatgaacgaa	atagacagat	cgctgagata	ggtcctcacc	3660
	tgattaaatc	ttggtaactg	tcagaccaag	tttacttcata	tataactttag	attgattttaa	3720
	aacttcattt	tttaataaaa	aggatctagg	tgaagatctt	ttttgataat	ctcatgacca	3780
	aaatccctta	acgtgagttt	tcgttccat	gaggcgtcaga	ccccgttagaa	aagatccaaag	3840
50	gatcttcttg	agatctttt	tttctgcgcg	taatctgtgt	tttgcaaaaca	aaaaaaccac	3900
	cgctaccagg	ggtggtttgc	tttgcggatc	aagagctacc	aacttctttt	ccgaaggtaa	3960
	ctggcttcag	cagagcgcag	ataccaaata	ctgtccttct	agtgtacccg	tagttaggcc	4020
	accacttcaa	gaactctgt	gcaccgccta	cataacctcg	tctgctaatc	ctgttaccag	4080
	tggctgctgc	cagtggcgat	aagtcgtgtc	ttaccgggtt	ggactcaaga	cgatagttac	4140
55	cggataaggc	gcagcggctcg	ggctgaacgg	gggttcgtg	cacacagccc	agcttggagc	4200
	gaacgaccta	cacccgaaactg	agataacctac	agcgtgagct	atgagaaagc	gcccacgtttc	4260
	cccgaaaggag	aaaggccggac	aggtatccgg	taagcggcg	ggtcggaaaca	ggagagcgc	4320
	cgagggagct	tccaggggg	aacgccttgcgt	atctttatag	tcctgtccgg	tttcgcccacc	4380
60	tctgacttgc	gctgtcgattt	ttgtgtatgt	cgtcagggggg	gcccggccata	tggaaaaaacg	4440
	coagcaacgc	ggccttttta	cggttccctgg	ccttttgcgt	gcctttgtct	cacatgttot	4500
	ttcctgcgtt	atcccctgtat	tctgtggata	accgtattac	cgcccttgag	tgagctgata	4560
	cgcgtcgccg	cagccgaaacg	accgagcgca	gctgacttgcgt	gagcgaggaa	gcccgaagagc	4620
	gcccaataacg	caaaccgcct	ctcccccgcgc	gttggccgt	tcattaaatgc	agctggcacc	4680
	acagggttcc	cgactggaaa	gccccggatgt	agcgtacccgc	aattaatgt	atgttagctca	4740
65	ctcatttaggc	accccaggtt	ttacacttta	tgttccggc	tcgtatgtt	tgtggaaattt	4800
	tgagcggata	acaatttca	acagggaaaca	gtatgacca	tgattaaatgc	aagctagccc	4860
	gggttagctt	gcatgcctgc	aggttt				4886

```

<210> 75
<211> 4905
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: vector
      pCMV-SSV(NNLS)

<400> 75
aaacagtccg atgtacgggc cagatatacg cgttgacatt gattattgac tagtattaa 60
tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
15 cttacggtaa atggccgcgc tggctgaccg cccaaacgacc cccgcccatt gacgtcaata 180
atgacgtatg ttcccatagt aacgcoata gggactttcc attgacgtca atgggtggac 240
tatttacggt aaactgcaca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgaccctt 360
tgggacttc ctacttggca gtacatctac gtatttagtca tcgcattttac catggtgatg 420
20 cgggttttgc agtacatcaa tgggcgtgga tagcggtttgc actcacgggg atttccaagt 480
ctccacccca ttgacgtcaa tggaggtttg ttttggcacc aaaatcaacg ggactttcca 540
aaatgtcgta aacaatccgc cccattgacg caaatggcgg gtaggcgtgt acgggtggag 600
gtctatataa gcagagctt ctggcttaact agagaaccca ctgcattactg gcttatacgaa 660
attaatacga ctcaactatag ggagacccaa gtcgtactcta gacttaattt agcgttgggg 720
25 tgagttactcc ctctcaaaag cgggcattgac ttctgcgtca agattgtcag tttccaaaaaa 780
cgaggaggat ttgatattca cctggccgc ggtgatgcct ttgaggggtgg cccgcgtccat 840
ctggtcagaa aagacaatct ttttgttgc aagctttaggg tttggcaggc ttgagatctg 900
gccatacact tgagtgcacat tgacatccac ttgcctttc tctccacagg tttccactcc 960
cagggcggcc gcaccatgcc caagaagaag aggaagggtga cgaaagataa gacgcgttat 1020
30 aaatacgggg attatatttt acgcgagagg aaagggcggt attatgttta caagctagag 1080
tatgaaaacg gtgaggtaaa agacgttac gttgggtcatt tagctgacgt cgttgaatca 1140
tatctaaaaa tgaatttagg ggtcgtaggg gataactccccc tacaagcggg tccccccgt 1200
ttcgagcccg ggacaagcgg aagcgggtgt ggaaaagagg gaactgaacg acgtaaaaata 1260
35 gcttgggttgc ccaatttgcg ccaatacgcg acggacggca acataaaggc gttctacaac 1320
tatctcatga acgaaaagggg gataagcggaa aaaactgcac aggactacat caatgctata 1380
tcaaaggcgt ataaagagac gagagacgcgca cagaaggctt accgactctt tgcacgtttc 1440
ttagcgtcac gcaatcatcat acatgtgaa ttgcggata aaatattgaa agcggtaaag 1500
gtgaagaagg cgaacgcgtg tatctcattt ccaacgttgg aagagataaa aaggacgtta 1560
40 caatttagca aagactatag cggaaaacgtc tacttcatctt accgtatcgc tctcgagtcg 1620
ggcgtaggc tgagcgaaat gcttagttgg actaggggat ataaggcgtt catttgcgt 1680
aacgacgtct gttattatcc gagagtagag gtacgttgc gggcaatagc cttctatgt 1740
ttccacataa cgcctctgaa cgacgtata aggacgttat gacgttgc aacttgcgc 1800
50 gctgagcttacgcgca aacattacgt gtttttacgc aatcgttacgc ggtttacaac ggtttacaac 1860
gccccatggc taaaagggtt ctgactcgag actggccgtc gttttacaac gtcgtactg gatccaggcg 1920
cgatcaata aaagatcatt atttcaata gatctgtgtt ttgggtttttt gtgtgcctt 1980
ggggaggggg agggcagaat gaggcgcggc caagggggag ggggaggcga aatgtttttt 2040
55 gggggaggggg gaggccagaa tgaccttggg ggagggggag gccagaatga gatccaggcg 2100
cggttacca gtcgaattt ctttcaataa gatctgtgtt ttgggtttttt gtgtgcctt 2160
ggcgttaccc aacttaatcg ctttcaataa gatctgtgtt ttgggtttttt gaatgacctt 2220
gaagaggccc gcaccgcgtc cccttcccaa cagttgcgcgca gctgtatgg ggcgcgcccc 2280
ctgtatgcgtt attttctctt ctcagttacaa tctgtctgaa gtcgtactt gatggcact 2340
55 gtcgtacgcgc cctgtacgggc ttgtctgtc coggtacgtc cttacagaca gctgtatgg 2400
gtctccggga gtcgtatgtt tcagagggtt tcaacgttcat caccgaaacg gtcgtatgg 2460
aagggcctcg tgatacgcct atttttagt gttaatgtca tgataataat gtcgtatgg 2520
acgtcagggtg gcaacttttcg gggaaatgtt cgcggaaaccc ctattttgtt gtcgtatgg 2580
atacatccaa atatgtatcc gtcgtatggaa caataacccct gataaaatgtc gccaacacccc 2640
60 tgaaaaaagga agatgtatgg tattcaacat ttccgtgtcg cccttattcc agctgtgacc 2700
gcattttggc ttctgtttt tgctcacca gaaacgcgtt tgaaatgtt ggtttcttag 2760
gatcagtgg gtcacgttgc gggttacatc gaactggatc tcaacagcgg attttctcaa 2820
gagagtttc gccccgaaga acgttttcca atgtatgacca cttttaaagt tcaataatat 2880
ggcgcggat tatccccgtat tgacgcggg caagagcaac tcgggtcgccg ctttttgcg 2940
65 tctcagaatg acttgggtga gtactcacca gtcacagaaa agcatcttac agatgctgaa 3000
acagtaagag aattatgcag tgctgcata accatgagtg ataacactgc tctgtctatgt 3060
cttctqacaa cgatcgagg accgaaggag ctaaccgctt ttttgcacaa catacaactat 3120
      ggtatggcatg ggcacactta 3180
      ggccaaactta 3240
      catgggggat 3300

```

5 catgtaaactc gccttgatcg ttgggaaccg gagotgaatg aagccatacc aaacgacgag 3360
 cgtgacacca cgatgcctgt agcaatggc acaacgttgc gcaaactatt aactggcga 3420
 ctacttactc tagcttcccg gcaacaatta atagacttga tggaggcga taaagttca 3480
 ggaccacttc tgcgtcggc ccttccggct ggctgggtta ttgctgataa atctggagcc 3540
 10 ggtgagcgtg ggtctcgccc tatcatttca gcactggggc cagatggtaa gcccctccgt 3600
 atcgttagtta tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagata 3660
 gctgagatag gtgcctcact gattaagcat tggtaactgt cagaccaagt ttactcatat 3720
 atactttaga ttgatttaaa acttcatttt taatttaaaa ggatcttagt gaagatcctt 3780
 tttgataatc tcatgaccaa aatccctaa cgtgagttt cgttccactg agcgtcagac 3840
 15 cccgtagaaa agatcaaagg atcttcttga gatctttt ttcgtcgcgt aatctgtc 3900
 ttgcaaacaa aaaaaccacc gctaccagcg gtgggttgtt tgccggatca agagctacca 3960
 actcttttc cgaaggtaac tggcttcagc agagcgcaga taccaaaatac tgiccttcta 4020
 gtgttagccgt agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgct 4080
 ctgctaattcc tggtaaccgt ggctgtcgc agtggcgata agtcgtgtct taccgggttg 4140
 20 gactcaagac gatagttacc ggataaggcg cagccgtcgg gctgaacggg gggttcgtgc 4200
 acacagccca gcttggagcg aacgaccatc accgaactga gataccatca gcggtgagcta 4260
 tgagaaaagcg ccacgttcc cgaagggaga aaggcgacca ggtatccggt aagcggcagg 4320
 gtcggaaacag gagagcgcac gaggggatc ccagggggaa acgcctggta tctttatagt 4380
 25 cctgtcgggt ttcgcccaccc tctgacttgcg cgtcgatttt tggatgcgc gtcagggggg 4440
 cggagcctat gggaaaacgc cagcaacgcg gccttttac gtttcgtgc cttttgttgg 4500
 cctttgctc acatgttctt tccctcgta tccctgtatt ctgtggataa cctgttattacc 4560
 qccttgagt gagctgatac cgctcgccg agccgaacga ccgagcgcag cgagtcagtg 4620
 agcgaggaag cggaaagagcg cccaaatacgc aaaccgcctc tccctcgccg ttggccgatt 4680
 30 cattaatgca gctggcacga cagggttccc gactggaaag cgggcagtga gcgcaacgc 4740
 attaatgtga gttagctcac tcattaggca ccccaggctt tacactttt gcttccgggt 4800
 cgtatgttgcgtt gttggaaatgtt gageggataa caatttcaca cagggaaacag ctatgaccat 4860
 gattacgcca agctagcccg ggctagcttgc catgcctgca gtttt 4905

30 <210> 76
 <211> 5290
 <212> DNA
 <213> Artificial Sequence

35 <220>
 <223> Description of Artificial Sequence: vector
 pCMVXisA

40 <400> 76
 agtccgatgt acgggccaga tatacgcgtt gacattgatt attgactagt tattaatagt 60
 aatcaattac ggggtcatta gttcatagcc catatatggc gttccgcgtt acataactta 120
 cggtaaatgg cccgcctggc tgaccgcaccc acgacccccc cccatttgcg tcaataatga 180
 cgtatgttcc catagtaacg ccaataggaa cttccatttgc acgtcaatgg gtggactatt 240
 45 tacggtaaac tggccacttgc gcaagtatc aagtgtatca tatgccaagt acgcccccta 300
 ttgacgtcaa tgacggtaaa tggccgcctt ggcattatgc ccagtacatg accttatggg 360
 actttcttac ttggcagttac atctacgtat tagtcatgcg tattaccatg gtgtatccgg 420
 ttggcagta catcaatggg cgtggatago ggtttgactc acggggattt ccaagtctcc 480
 accccatttgc cgtcaatggg agttttttt ggcacaaaaa tcaacgggac tttccaaaat 540
 50 gtcgtacaaa ctccggccca ttggacggaaa tggccgttag gctgtacgg tgggagggtct 600
 atataaaggcag agctctctgg ctaacttagag aaccacttgc ttactggctt atcggaaattta 660
 atacgactca ctataggggg acccaagctg actctagact taattaagcg ttgggggttag 720
 tactccctct caaaagcgccg catgacttct ggcctaagat tgcgttttc caaaaacgag 780
 gaggatttgc tattcacctg gcccgcgtt atgcctttga gggtggccgc gtcctatgtt 840
 55 tcacacttgcg tgacatttgc atccactttt cctttctctc cacagggtgc cactcccagg 960
 gccccccccc gatatgcggg atcagggtca agacaaatat caacaaggct ttgcagactt 1020
 agagccactc tcatctaccg acggcaggttt tctcggtc agtctgcaag cacagcagca 1080
 aagagaacac atgagaacaa aagtactaca agaccttagac aaggtaaattc tgcgtttgaa 1140
 60 gtcgtcaaaac acgaaagtct cagttcgaga atctaacggc agtctgcaat tacgagcaac 1200
 gttaccaattt aaacctggag ataaggacac caacggtaca ggcagaaaagc aatacaatct 1260
 cagcttgcattt atccctgcaacttgg actgaagacg gctgaggaaag aagctttaga 1320
 attagttaaa ttaatcgctc gggaaaactt tgaatggaaat gataaatatt taggcaaaaga 1380
 agccactaaa aaagatttgc acacaaatagg tgatttacta gaaaatttg cagaagagta 1440
 tttttaaaacc catabaacgc ccaactaaag cgaacatacc ttttttactt attttcccg 1500
 65 caccacacgata tataccaattt ccaaaagattt agcaacggcg gaaaatctca tcaattcaat 1560
 ttagcaaaatc gataaagaat gggcgagata taatgcgcgc agagccatatt cagcttttg 1620
 cataacatttca aatatacgaaa ttgatttgc ccagtttcc aaaaatgccttgc atcgcaattc 1680

5	ggcgcacatc	cccacacatg	cagaataact	atcaggaaatt	accaaatttg	aagactatct	1740
	agtttaccaga	ggaaatcaag	ttaatgaaga	tgtaaaagat	agctggcaac	tttggcgctg	1800
	gacatatgga	atgttagcag	tttttggttt	acgccccagg	gaaatttttta	ttaaccctaa	1860
	tattgattgg	tggtaagca	aagagaataat	agacccatcaca	tggaaagttag	acaaagaatg	1920
	taaaaactgg	gaaagacaag	cattaccctt	acataaaagaa	tggattgatg	agtttgattt	1980
	aagaaatccg	aaatattttag	aatatgtggc	aacagcaattt	agtaaaaaaag	ataaaacaaa	2040
10	tcatgtgaa	ataacagccat	taactcagcg	tatttagttgg	tggtttcgga	aagtgcattt	2100
	agattttaaa	ccctatgatt	tagtccacgc	ctgggcaattt	agagcgcata	ttttaggcat	2160
	accaatcaa	gcggcggctg	ataatttggg	gcatagttatg	caggttcata	cacaacacta	2220
	tcagcgctgg	ttctcgctag	atatgcggaa	gttagcgatt	aatcaggctt	tgactaaagag	2280
	gaatgaattt	gaggtgatta	gggaggagaa	tgctaaattt	cagatagaaa	atgaaaggtt	2340
	gaggatggaa	attgagaagt	taaagatgg	aatagcttat	aagaatagt	gagcggccgc	2400
15	gttgcaccc	agatccaggc	gcccataat	aaaagatcat	tattttcaat	agatctgtgt	2460
	gttgggtttt	tgtgtgcctt	gggggggggg	gaggccagaa	tgaggcgccg	ccaaggggga	2520
	ggggggggcc	agaatgcct	tgggggggg	ggggccaga	atgaccctgg	gggaggggga	2580
	ggccagaatg	aggcgccccc	cgggttaccc	agctcgaattt	cactggccgt	cgttttacaa	2640
	cgtctgtgat	ggggaaaaccc	tggcgttacc	caacttaatc	gccttgcagc	acatccccct	2700
	ttagccagat	ggcgtaatag	cgaagaggcc	cgaccggatc	gcccttccca	acagttggcgc	2760
20	agccctgaatg	gccaatggcg	cctgatgcgg	tattttctcc	ttacgcattt	gtgcggattt	2820
	tcacaccgca	tatggtgcac	tctcagttaca	atctgcctgt	atgcccata	gttaagccag	2880
	ccccgacacc	cgccaaacacc	cgctgaogcg	ccctgacggg	cttgcgttgc	cccgccatcc	2940
	gcttacagac	aagctgtgac	cgtctccggg	agctgcattgt	gtcagagggtt	ttcacccgtca	3000
	tcaccgaaac	gcccggagacg	aaagggctc	gtgatacggcc	tatttttata	ggtaatgtc	3060
	atgataataa	tgggttctta	gacgtcagg	ggcactttt	ggggaaatgt	gcccggaaacc	3120
25	cctattttgtt	tatttttctta	aatacattca	aatatgtatc	cgctcatgag	acaataaccc	3180
	tgataaaatgc	ttcaataata	ttgaaaaagg	aagagttatg	gtattcaaca	tttccgtgtc	3240
	gcccttattt	ccttttttgc	ggcattttgc	cttcctgttt	ttgcacccat	agaaacgttg	3300
	gtgaaagtaa	aagatgtcga	agatcaattt	ttgtgcacgg	tgggttacat	cgaactggat	3360
30	ctcaacagcg	gtaagatccct	tgagagttt	cgccccggaa	aacgttttcc	aatgtatgagc	3420
	acttttaaag	ttctgtctatg	ttggcgggtt	ttatcccgta	ttgacgcccgg	gcaagagcaa	3480
	ctcggtcggc	gcatacacta	ttctcagaat	gacttggttt	agtactcacc	agtcacagaa	3540
	aagcatctt	cggtatggat	gacagtaaga	gaattatgca	gtgctccat	aaccatgagt	3600
	gataaactcg	cggccaaactt	acttctgaca	agatcggag	gaccggagga	gctaaccgc	3660
35	tttttgcaca	acatggggga	tcatgttaat	cgcccttgcac	gttgggaaacc	ggagctgaat	3720
	gaaggccatac	caaaccgcga	gcgtgacacc	acgatgcctg	tagcaatggc	aacaacgttg	3780
	cggaaactat	taacttggcg	actacttaat	ctagttcccc	ggcaacaattt	aatagactgg	3840
	atggaggccg	ataaaagttgc	aggaccactt	ctgctgcctgg	cccttccggc	tggctgggtt	3900
	attgtctgata	aatctggagc	cggtgaggct	gggtctcgcg	gtatcattgc	agcaactggg	3960
40	ccagatggta	agccctcccg	tatcgtagtt	atctacacga	cgggggagtca	ggcaactatg	4020
	gatgaacgaa	atagacagat	cgctgagata	ggtgcctcac	tgattaaagca	ttggtaactg	4080
	tcagaccaag	tttactcata	tatactttt	attgattttaa	aacttcattt	ttaattttaa	4140
	aggatctagg	tgaagatccct	ttttgataat	ctcatgacca	aaatccctta	acgtgagttt	4200
	tcgttccact	gagcgtcaga	ccccgtagaa	aagatcaaag	gatcttcttg	agatcccttt	4260
45	tttctgcgct	taatctgtgt	cttgcaaaaca	aaaaaaaccac	cgctaccagc	gggtgtttgt	4320
	ttggccggatc	aagagctacc	aactctttt	ccgaaggtaa	ctggcttcag	cagagcgcag	4380
	ataccaaata	ctgtcccttc	agtgtaccc	tagttggcc	accacttcaa	gaactctgt	4440
	gcaccgccta	catacctcgc	tctgtctatc	ctgttaccag	tggctgtctgc	cagttggcgat	4500
	aagtctgtc	ttaccgggtt	ggactcaaga	cgatagttac	cggataaggc	gcagcggctg	4560
50	ggctgaacgg	ggggttctgt	cacacagccc	acgtttggagc	gaacgaccta	caccgaactg	4620
	agataccatc	agcgtgagct	atgagaaagc	gocacgcttc	ccgaaggggag	aaaggccggac	4680
	aggatcccg	taagcggcag	ggtcggaaaca	ggagagcgc	cgagggagct	tccaggggga	4740
	aacgcctgtt	atctttatag	taactgtcggt	tttcgcacc	tctgacttgc	gcgtcgattt	4800
	ttgtgtatgt	cgtcagggggg	ggggggccta	tggaaaaacgg	ccagcaacgc	ggccctttta	4860
55	cggttccctgg	ccttttgcgt	gcctttgtct	cacatgttct	ttcctgcgtt	atcccctgtat	4920
	tctgtggata	accgttattac	cgectttag	tgagctgata	ccgctcggcg	cagccgaacg	4980
	accgagcgca	gcgagtcagt	gagcaggaa	ggcgaagagc	gccccataac	caaaccgcct	5040
	cgtccccgcgc	gttggccgtat	tcattaaatgc	agctggcact	acaggtttcc	cgactggaaa	5100
	ggggggcagt	agcgcacacgc	aattaaatgt	agttagctca	ctcatttaggc	accccaggt	5160
60	ttacacttta	tgcttcggc	tgtatgttgc	tgtggaaattt	tgagcggata	acaattttcac	5220
	acaggaaaaca	gctatgacca	tgattacgccc	aagctagccc	gggctagctt	gcatgcctgc	5280
	aggtttaaac						5290

65 <210> 77
<211> 5309
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
pCMVXisANNLS

5 gaggaccgaa ggagctaacc gctttttgc acaaactggg ggatcatgta actgccttg 3720
 atcggtggaa accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc 3780
 ctgttagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagct 3840
 cccggcaaca attaatagac tggatggagg cggataaaagt tgcaggacca cttctgcgt 3900
 10 cggcccttcc ggctggctgg ttattgtctg ataaatctgg agccggtagg cgtgggtctc 3960
 gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca 4020
 cgacggggag tcagcactt atggatgaaac gaaatagaca gatgcgttag ataggtgcct 4080
 cactgattaa gcattggtaa ctgtcagacc aagttaactc atataactt tagattgatt 4140
 15 taaaacttca ttittaaatit aaaaggatct aggtgaagat cctttttgtat aatctcatga 4200
 caaaaatccc ttaacgttag ttttcgtcc actgagcgtc agaccccgta gaaaagatca 4260
 aaggatcttc ttgagatct ttttttgcgctc gcttaatctg ctgcttgcgaa aaaaaaaaaac 4320
 caccgatacc agcgggtggtt tggttgcggg atcaagagct accaactctt ttccgaagg 4380
 taactggctt cagcagagcg cagataccaa atactgtctt tctagtgttag ccgtagttag 4440
 20 gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac 4500
 cagtggctgc tgccagtggc gataagtctg gtcttaccgg gttggactca agacgatagt 4560
 taccggataa ggcgcagcgg agcgaacgac ctacaccgaa ttcccgaaagg gacagggtatc 4620
 25 gacagggtatc cggtaagcgg cagggtcgg acaggagagc 4740
 gcacgagggaa gcttccaggg gaaaacgcct ggtatctta tagtccgtc ggtttcggc 4800
 acctctgact tgagcgtcga tttttgtat gctcgtcagg gggggggagc ctatggaaaa 4860
 acgcccagcaa cgcggccctt ttacggttcc tggcctttt ctggcctttt gctcacatgt 4920
 tctttcttgc gttatcccct gattctgtgg ataaaccgtat taccgccttt gagtgagctg 4980
 ataccgctcg cgcagccga acgaccgagc gcaagcgtac agttagcggag 5040
 30 agcgcctaatt acgcaaaaccg acgacaggtt tccggactgg 5100
 acgacatggggccat tcaacttataa ggcaccccccag ttgtgagcggg 5160
 ttgtgagcggg ataaacaattt gtttacact ttatgcctcc ggtctgtatg 5220
 cccgggatcgatgccc ttcagggttt 5280
 35 <210> 78
 <211> 7608
 <212> DNA
 <213> Artificial Sequence
 <220>
 40 <223> Description of Artificial Sequence: vector
 pPGKnifD
 <400> 78
 tcgaggaatt ctaccggta ggggaggcgc ttttcccaag gcagtcgttgc gcatgcgtt 60
 tagcaggcccc gctgggcact tggcgttaca caagtggctt ctggcgtcga cacattccac 120
 atccccccgtt aggcccaac cggctccgtt ctttggcgc cccttcgcgc caccttctac 180
 45 tcctcccttca gtcaggaagt tcccccccgcc cccgcagtc ggtctgtcga ggacgtgaca 240
 aatggaaatgta gcacgttca ctatgtctgttgc gcaatggac agcaccgcgt agcaatggaa 300
 gcggttaggc ctttggggca ggggccaataa gcaatgttgc tccttcgtt tctgggctca 360
 gaggctggaa aggggtgggt cggggggcgg gctcaggggc gggctcagggg gcggggcggg 420
 50 cgcccgaaagg tcctccggag gccccggatt ctgcacgtt caaaagcgtca cgtctgcgc 480
 gctgttctcc ttttccat tccgggctt ttcgacactgc agccgggtac agttcgaaatg 540
 gcttccctt tccgtcaaat gcaactcttgc gattacttgc aaccttgcga tggggtgcaa 600
 atgtcagatc agataagttc gaataacttc gtatagcata cattatacgta agttataagc 660
 ttgcattgttgc ttcggatccgc gcccacgacc ggcggccgg tgccgcacc atccccctgac 720
 ccacggccctt gacccttacac aaggagacga ctttccatga ccgagtacaa gcccacgggt 780
 55 cgccctcgcca ccccgacgac gttccccgg gccgtacgca ccctcgccgc cgcgttgcggc 840
 gactaccccg ccacgcgcca caccgtcgac ccggaccgc acatcgagcg ggtcaccggag 900
 ctgcagaagaaat ttttccatca cgcgttgcgg ctgcacatog gcaagggtgtt ggtcggggac 960
 gacggcgcgcg cgggtggcggt tggaccacac ccggagagcgc tcgaaggcggg ggcgggtgtt 1020
 60 gccgagatcg gcccgcgcatt ggcggagggtt aaggagcccg cgtggttctt ggcacccgtc 1080
 atggaggccc ttctggcgcc gcaacggggcc aaggagcccg cgtggttctt ggcacccgtc 1140
 ggcgttctcgcc cgcacccatca gggcaagggtt ctggcgtcgc cgcgttctt ggcacccgtc 1200
 gaggcggccg agcgcggccgg ggtcccccccttccatggaga cttccgcgc acgcacccatca 1260
 cccttctacg agcggctcggtt ttcacccgtt acggccgacg tcgagtgcggc gaggagccgc 1320
 gcgacctgtt gcatgaccccg caagcccggtt ggcgtacgac cgcggccacgac cccgcacgc 1380
 65 cgcacccgaa ggagcgcacg accccatggc tccgaccgaa gccgaccggc gcgccccccgc 1440
 cgaccccgca cccggcccccgg agggccacccg actcttagagg atcataatca gccataccac 1500
 attttagagat gtttacttgc taaaatggat gcaattgttgc tttttaatctt gtttattgca gtttataatg 1560
 taatgtatggat gtttacttgc tttttaatctt gtttattgca gtttataatg 1620

5	aagcaatacg atcacaatt tcacaaataa agcattttt tcactgcatt ctatgtgg 1680 tttgtccaaa ctcataatg tatcttatca tgtctggatc cagctgtga aagctattaa 1740 accacaaaaa ggattactcc ggccttatac acggtaacga cggatttggc tccataactt 1800 cgtatagcat acattatacg aagttatacc gggccaccat ggtcgcgagt agcttggcac 1860 tggcgtcgt tttacaacgt cgtactggg aaaaccctgg cgtaacccaa cttaatcgcc 1920 ttgcagcaca tcccccttgc gccagctggc gtaatagcga agaggcccgc accgatcgcc 1980 cttcccaaca gttgogcagc ctgaatggcg aatggcgctt tgccctgggtt cccgcaccag 2040 aagcggtgcc ggaaagctgg ctggagtgcc atcttcttgc gcccgcatact gtcgtcgcc 2100 cctcaaactg gcagatgcac gtttacatgc ccccatctt caccacgtt acctatccca 2160 10 ttacggtcaa tccggcggtt gttccacgg agaattccgac ggggtttaac tcgctcacat 2220 ttaatgttgta taaaatgtgg ctacaggaa gccagacgac aattttttt gatggcgta 2280 actcggcgtt tcatctgtgg tgccgtctga atttgcacgtt gtcacccggc gacagtcgtt 2340 ttacggtgcg ttggagtgtac agcgcatttt tacgcgcggg agaaaaccgc ctcgcggta 2400 gcattttccg tgacgtctcg ggcagttaatc tggaaagatca ggatatgtgg cggatgaggc 2460 ttgctgcata aaccgactac acaaattcago gatttccatg 2520 ttgcactctcg cttaatgtat gattcagcc ggcgtgtact ggaggctgaa gttcagatgt 2580 gccccgagtt gcgtgactac ctacggtaa cagtttctt atggcagggt gaaacgcagg 2640 tcgcccagcgg caccgcgcct ttcggcggtt aaattatcga tgagctgtgt gttatgcgg 2700 atcgcgtcac actacgtctg aacgtcgaaa acccgaaact gtggagcgc gaaatcccg 2760 20 atctctatcg tgcgtgttt gaaactgcaca cccgcgacgg cacgctgatt gaagcagaag 2820 cctgcgtatgt cggttccgc gaggtgcgga ttgaaaatgg tctgtctgt ctgaacggc 2880 agccgttgc gattcgaggc gttAACCGTC acgagcatca tccctctgcatt ggtcaggta 2940 ttggatgagca gacgatgtgtt caggatatcc tgcgtatgaa gcagaacaac tttAACCGCC 3000 tgcgtgttc gcatatccg aaccatccgc tgggttacac gctgtgcac cgctacggcc 3060 25 tggatgtgtt ggtatgtggg aatattggaa cccacggcat ggtgcacatg aatcgctctga 3120 cccgatgtcc ggcgtggcta atcgatataa cccggatgtt atggcagggt gaaacgcagg 3180 atcgatataa cccgagtgtt atcatctgtt cgtgtggaa tgaatcaggc caccggcgta 3240 atcacgacgc gctgtatcg tggatcaat ctgcgtatcc ttccggccg gtgcgtatg 3300 aaggcggcgg agccgacacc acggccaccg atattatttgc cccgatgtac ggcgcgtgg 3360 30 atgaagacca gcccctcccg ctggagagac ggcggcgtt gtcgtatcc aatcgccca 3420 gccccggcgtt taaatactgg tctggactg ggtggatcag ctgcgtatcc aatcgccca 3480 tttacggcgg tgatttggc tcttgcgaa cccggatgtt aatcgccca cccgatgggt 3540 35 tcttgcgaa cccgacggc tccagttccg tttatccggg gtcgtatcc aatcgccca 3600 gcatccggcgtt gtcgtatcc aatcgccca cccgatgggt aacagtcttgc 3660 aaggcggcgg aatcgccca cccgatgttgc tgcgtatcc aatcgccca 3720 atcgatataa cccgatgttgc ttcacccgtt aatcgccca cccgatgttgc 3780 40 catggtcaga agccggcgcgtt gtcgtatcc aatcgccca 3840 gtgtgacgct cccggccgcgtt gtcgtatcc aatcgccca 3900 tttgcgtatcc gtcgtggtaat aacaggatgtt gtcgtatcc aatcgccca 3960 agatgtggat tggcataaa aacaggatgtt gtcgtatcc aatcgccca 4020 45 caccgctgta taacgacatt tccacccgtt gtcgtatcc aatcgccca 4080 tcgaacgctg gaaggcggcgtt gtcgtatcc aatcgccca 4140 cagatacact tgcgtatcc aatcgccca 4200 aaaccttatt ttcacccgtt gtcgtatcc aatcgccca 4260 ccgttgcgtt tgaagtggcgtt gtcgtatcc aatcgccca 4320 50 agctggcgtt ggtggatcag gtcgtatcc aatcgccca 4380 ccgaccgcct tactggccgcgtt gtcgtatcc aatcgccca 4440 ccccgtacgt ttcacccgtt gtcgtatcc aatcgccca 4500 gcccacacca gtggcgccgcgtt gtcgtatcc aatcgccca 4560 tgatggaaac cagccatcgatcc gtcgtatcc aatcgccca 4620 acggtttcca tatggggatt gtcgtatcc aatcgccca 4680 55 tccagctgag cccggcgtcgcgtt gtcgtatcc aatcgccca 4740 cgggcagggg ggatctttgt ttcacccgtt gtcgtatcc aatcgccca 4800 ctacctacag agatttaaag aacacatccgtt gtcgtatcc aatcgccca 4860 aaactactgtt ttcttaattgtt gtcgtatcc aatcgccca 4920 gagcagctgtt ggaatgcgcgtt gtcgtatcc aatcgccca 4980 60 cacaactaga atgcgtatccgtt gtcgtatcc aatcgccca 5040 atttgcatacc attataatgg ttcacccgtt gtcgtatcc aatcgccca 5100 gtttcagggtt cagggggagg ttcacccgtt gtcgtatcc aatcgccca 5160 ttgttatggctt gattatgcgtt gtcgtatcc aatcgccca 5220 taatgtcatgtt aataataatggt gtcgtatcc aatcgccca 5280 65 cggaaaccctt atttgcgtatccgtt gtcgtatcc aatcgccca 5340 ataaccctgatccgtt ttttgcgtatccgtt gtcgtatcc aatcgccca 5400 ccgtgtcgccctt ttttgcgtatccgtt gtcgtatcc aatcgccca 5460 ttttcttaat acattcaat atgtatccgc tcatgagaca 5520 aataatattgtt aaaaaggaag agtatgagta ttcaacatttgcgtatcc 5580 ccgtgtcgccctt ttttgcgtatccgtt gtcgtatcc aatcgccca 5640
---	---

aacgctggtg aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg gttacatcga 5700
 actggatctc aacagcgtta agatccttga gagttttcgc cccgaagaac gttttccaaat 5760
 gatgagcaact tttaaagttc tgctatgtgg cgcgttatta tcccgatattt acgcggggca 5820
 5 agagcaactc ggtcgccgca tacactattc tcagaatgac ttggttgagt actcaccatg 5880
 cacagaaaag catcttacgg atggcatgac agtaagagaa ttatgcgtg ctgcataaac 5940
 catgagtgt aacactcggtt ccaacttact tctgacaacg atcggaggac cgaaggagct 6000
 aaccgctttt ttgcacaaca tggggatca tggtaactcgc cttgatcggtt gggaaaccgg 6060
 gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgttag caatggcaac 6120
 aacgttgcgc aaactattaa ctggcgaact attactcta gtttccggc aacaattaat 6180
 10 agactggatg gaggcggata aagttgcagg accattctg cgtcgcccc ttccggctgg 6240
 ctggtttattt gctgataaat ctggagccgg tgagcgtggg tctcgccgtt tcaattgcagc 6300
 actggggcca gatggtaagc cctcccgat ctagttatc tacacgacgg ggagtccaggc 6360
 aactatggat gaacgaaata gacagatcgc tggataggt gcctactga ttaagcattt 6420
 gtaactgtca gaccaagttt actcatatat acttagatt gattaaaac ttcatttttt 6480
 15 attaaaagg atcttaggtga agatccttt tgataatctc atgacaaaaa tcccttaacg 6540
 tgagtttgc ttccactgag cgtcagaccc cgtagaaaag atcaaaaggat ctcttigaga 6600
 tcctttttt ctgcgttaa tctgtgtttt gcaaaacaaaaa aaaccaccgc taccagcggt 6660
 gttttgtttt ccggatcaag agctaccaac tcttttccg aaggtactg gttcagcag 6720
 20 agcgcagata ccaaatactg tcctttagt gtagccgtag ttaggccacc acttcaagaa 6780
 ctctgttagca ccgcctacat acctcgctc gctaattctg ttaccagtgg ctgctgccc 6840
 tgccgataag tcgtgtctt cgggttttga ctcagacga tagttaccgg ataaggcgc 6900
 gccgtcgggc tgaacggggg gttcgtgcac acagcccacg ttggagcga cgcctacac 6960
 cgaactgaga tacctacgc gtgagctatg agaaagcggc acgctttccg aagggagaaa 7020
 ggcggacagg tatccggtaa gcggcagggt cggAACAGGAA 7080
 25 agggggaaac gcctgttac ttatagttc tgcgggtttt cgcacactt gacttgagcg 7140
 tcgatttttt tgatgtctgtt cagggggggc gaggctatgg aaaaacgcca gcaacgcggc 7200
 cttttacgg ttccctggct tttgctggcc ttttgcac atgttcttc ctgcgttattc 7260
 ccctgattct gtggataacc gtattacccg ctttgagtga gctgataacc ctgcggccag 7320
 30 ccgaacgacc gagegcagcg agtcagtgag cgaggaagcg gaagagcgc caataacgca 7380
 accgccttc cccgcgcgtt ggcgcgatca ttaatgcagc tggcacgaca gttttcccg 7440
 ctggaaagcg ggcagtgago gcaacgcaat taatgtgagt tagctcaactc attaggcacc 7500
 ccaggctta cacttatgc ttccggctcg tatgttgtt ggaattgtga gcccataaca 7560
 atttcacaca ggaaacagct atgaccatga ttacgccaag ctggcgcg 7608

35 <210> 79
 <211> 7523
 <212> DNA
 <213> Artificial Sequence

40 <220>
 <223> Description of Artificial Sequence: vector
 pPGKnifD3'

45 <400> 79
 tcgaggaatt ctaccggta ggggaggcgc ttttccaaag gcagtcgttgc gcatgcgtt 60
 tagcagcccc gctgggact tggagctaca caagtggctt ctggctcgca cacattccac 120
 atccaccgggt aggcgcacac cggctccgtt ctttgggtgc cccttcgcgc cacattctac 180
 toctccctta gtcaggaatg tccccccgc cccgcagctc gctcgttgc gacgtgaca 240
 50 aatggaaatg gcacgtctca ctgttcgtt gcaatggac agcaccgtg agcaatggaa 300
 gccggtaggc ctttggggca gggccaaata gcaacttgc tccttcgtt tctgggttca 360
 gaggctgggaa aggggtgggt ccggggggcgg gctcaggggc gggctcaggg gccggggggg 420
 cggccgcagg tcctccggag gcccggcatt ctgcacgtt caaaacgcga cgtctgcgc 480
 gctgttctcc tcttcctcat ctccgggtt ttcgacactgc agcccggtac agttcgaata 540
 55 acttcgtata gcatacatta tacaaggta taagcttgc tgcctgcagg tggccgc 600
 cggccggccg gccgtgtccg ccaccatccc ctgacccacg cccctgacc ctcacaagg 660
 gacgacccctt catgaccggag tacaaggcca cgggtgcgcct cggccacccgc gacgacgtcc 720
 cccggccgtt acggccaccctt gcccggcgt tccggacta cccggccacccg cggccacacc 780
 tccgacccggaa cggccacatc gagggggtca cggagctgca agaacttgc ctcacgcgc 840
 60 tccggcttgcgatc cttccggatgg gtcgtgggtcg cggacgacgg cggccgggtt gccatggcc 900
 ccacgcggaa gagcgtcgaa gggggggcgg tggccggatgg gatcggcccg cgcacatggcc 960
 agttgagccgg ttccggctgtt gccgcgcggc aacagatggaa aggcccttcg ggcggccacc 1020
 ggcccaaggaa gccccgtgg ttccctggca cctgcggcgtt ctcggccgc caccagggg 1080
 65 agggtctggg cagccgcgtt gtcgtccccc gatggagggc gcccggccg gccgggggtgc 1140
 cccgccttcctt ggagacccctt gccggccggc acctccctt ctacgagcgg ctcggcttca 1200
 cccgtcaccgc cggccgttgc gccggccggc accggccgcac ctgggtcgatg accccgcaagc 1260
 cccggccgttgc acggccggcc caccggccgc agccggccac cggccggcc 1320

atggctccga ccgaagccga cccgggcggc ccccccggacc ccgcacccgc ccccgaggcc 1380
 caccgactct agaggatcat aatcagccat accacatttg tagaggtttt acttgcttta 1440
 aaaaacctcc cacacctccc cctgaacctg aaacataaaa tgaatgcaat tttgttgtt 1500
 5 aacttggtttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 1560
 aataaagcat ttttttact gcattctagt tttgggtttgt ccaaactcat caatgtatct 1620
 tatcatgtct ggatccagct gttgaaagct attaaaccac aaaaaggatt actccggccc 1680
 ttatcacggt tacgacggat ttggatccat aacttcgtat agcatacatt atacgaagtt 1740
 ataccgggccc accatggctcg cgagtagctt ggcactggcc 1800
 10 ctggggaaaac cctggcggtt cccaaactaa tcgccttgcga gcacatcccc cttecgccag 1860
 ctggcgtaat agcgaagagg ccgcacccga tcgccttcc caacaggatgc gcaagctgaa 1920
 tggcgaatgg cgcttgcct ggtttccggc accagaagcg gttgggaaa gctggctgga 1980
 gtgcgatctt cctgaggccg atactgtcgt cgccccctca aactggcaga tgcacgggtt 2040
 15 cgatgcgcctt atctacacca acgtaaaccta tcccattacg gtcacatccgc cgtttgttcc 2100
 cacggagaat ccgacgggtt gttactcgct cacatttaat gttgatgaaa gctggctaca 2160
 ggaaggccag acgcaatta tttttgatgg cgtaactcg gctttcatac tttgttgca 2220
 cgggcgctgg gtcgggttacg gccaggacag tcgtttgcgg tctgaatttg acctgagcgc 2280
 attttacgc gccggagaaa accgcctcgc ggtgatgggt 2340
 ttatctggaa gatcaggata tttggcggat gagcggcatt ttccgtgacg ttcgttgc 2400
 20 gcataaaaccg actacacaaa tcagcgattt ccatgttgcg actcgcttta atgatgattt 2460
 cagccgcgt gtactggagg ctgaagtca gatgtgcggc gagttgcgtg actacactac 2520
 ggttaacagtt tctttatggc agggtgaaac gcagggtcgcg agcggcaccg cgccttcgg 2580
 cggtgaaattt atcgatgagc gtgggttta tgccgatcgc gtcacactac gtcgttgc 2640
 25 cggaaaaccgg aaactgtgga gcccggaaat cccgaatctc ttcgttgcg tttgttgc 2700
 gcaacaccggc gacggcacgc tgattgaagc agaaggctgc gatgtcggtt tcgcgcagg 2760
 gcggtattgaa aatggctcgc tgctgtcgaa cggcaagccg ttgtgtattt gaggcggtt 2820
 cccgtacggc catcatcctc tgcatggta ggtcatggat gacgacgca gttgtc 2880
 tatectgtc atgaagcaga acaactttaa cgcgtgcgc ttttcgttgcatt atccgaacca 2940
 30 tccgtgttgc tacacgctgt ggcgaccgtt cggcgttat gttgtgttgc aacccaaat 3000
 tggaaaccac ggcgttgcgca caatgaatcg ttttgcgtat gatccgcgtt gggttccggc 3060
 gatgagcggaa cgcgttacgc gaatggtca ggcgcgttac aatcaccggc gttgtatcat 3120
 ctggtcgttgc gggaatgat caggccacgg cgcttaatcac gacgcgttgcgtt atcgttgc 3180
 35 caaatctgtc gatccttccc gcccgggtca gtatgaaggc ggcggagccg acaccacggc 3240
 caccgatatt atttgcggc gttacgcgcg ctttggatgaa gaccaggccct tcccggttgc 3300
 gccgaaatgg tccatcaaaa aatggcttc gctacctggc gagacgcgcg cgctgtatcc 3360
 ttgcgaatac gcccacgcg tggtaacag ttttggcggt ttcgtttaat atggcaggc 3420
 40 gtttcgtcag tatccccgtt tacaggccgg cttcgttgcg gactgggtgg atcgtcgct 3480
 gattaaatat gatggaaaacg gcaaccctg gtcgggttac ggcgggttgcattt ttggcgat 3540
 gccgaacgat cgccaggatct gtatgaacgg ttttggcggtt gccgaccgc cggccatcc 3600
 agcgctgacg gaagaaaaac accagcagca gttttccag ttcgtttat ccgggaaac 3660
 45 catcgaaatg accagcgaat acctgttccg tcatagcgat aacgagctcc tgcactggat 3720
 ggtggcgctg gatgttgcg cgttgcgaag cggtaagtg cctctggatg tgcgtccaca 3780
 aggttaacacg ttgttgcgac tgcctgaaact accgcagccg gagagcgcgc ggcactctg 3840
 gtcacagta cgcgttagtc aaccgaacgc gaccgcgttgc tcagaacgc ggcacatcag 3900
 50 cgcctggcgcatgttgcgcatggc gtttggggaaa ctttgcgttgc acgctccccc cgcgtcccc 3960
 cgcctcccg catctgacca ccagcgaaat ggatttttgc atcgagctgg gtaataagcg 4020
 ttggcaattt aaccggcagt caggcttct ttcacagatg tggattggcg ataaaaaaca 4080
 actgctgacg ccgtcgccg atcagttcac ccgtgcaccg ctggataacg acattggcg 4140
 aagtgaagcg accccgattt acccttaacgc ctgggtcgaa cgctggaaagg cggcggggcca 4200
 ttaccaggcc gaagcagcgt ttttgcgttgc caccggcagat acacttgcgt attcgttgc 4260
 55 gattacgacc gtcacgcgt ggcagcatca gggaaaaacc ttatttatca gccggaaaac 4320
 ctaccggatt gatgttagt gtcacaaatggc gattaccgtt gatgttgaag tggcgagcga 4380
 tacaccgcattt ccggcgccg ttggcctgaa ctggcagctg ggcaggttag cagagcgggt 4440
 aaacttgcgc ggtttagggc cgcaagaaaa ctatccccgc cgccttactg cccctgttt 4500
 tgaccgttgc gatctgccc ttttgcgttgc gtcagacat gtatccccgc tacgttcc cggcgaaaa 4560
 55 cggctcgcc tggggacgc gogaatttggaa ttatggccca caccagtggc gccggcactt 4620
 ccagttcaac atcgcgcgt acagtcaaca gcaactgtat gaaaccaagcc atcgcctatct 4680
 gtcgcacgcg gaagaaggca catggctgaa ttttgcgttgc ttccatatgg gatgttgc 4740
 cgacgacttc tggggccgt ctttgcgttgc ggaatttccag ctggcgccg gtcgttacc 4800
 60 ttaccaggcc ttgttgcgttgc aaaaataata ataaaccggc agggggatc ttgttgc 4860
 aaccttactt ctgttgcgttgc acataatgg acaaactacc tacagagatt taaagcttca 4920
 aggttaaatat aaaatttttt agtgtataat gttttgcgttgc actgatcttca atttttgtt 4980
 tatttttagt tccacacccat ggaactgtat aatgggagca gttgttgc gttgttgc 5040
 gacatgataa gatacattga tgatgttgc ctttgcgttgc ttttgcgttgc gttgttgc 5100
 tgcttttattt gtgaaattttt tgatgttgc gtttgcgttgc ttttgcgttgc ttttgcgttgc 5160
 65 aaacaagttt acaacaacaa ttgcattcat ttttgcgttgc agtttgcgttgc gttgttgc 5220
 gaggtttttt aaagcaagttt aaaccttac aatgtggta ttttgcgttgc ttttgcgttgc 5280
 ccgcaggggcc tcgtgatacg ctttgcgttgc tagttaatg toatgataat aatggtttct 5340

5	tagacgtcag	gtggcacttt	tccggggaaat	gtgcgcggaa	ccccttatttgc	tttatttttttc	5400
	taaatacatt	caaatatgtt	tccgctcatg	agacaataac	cctgataaat	gcttcaataa	5460
	tattgaaaaa	ggaagagttat	gagtattcaa	catttccgtg	tcgcccattt	tccctttttt	5520
	gcggcatttt	gccttcctgt	tttigctcac	ccagaaacgc	tggtggaaagt	aaaagatgt	5580
	gaagatca	tgggtgcacg	agtgggttac	atcgaactgg	atctcaacag	cggtaagatc	5640
	cttggaggtt	ttccccccga	agaacgtttt	ccaatgtga	gcactttttaa	agttctgcta	5700
	tgtggcgcgg	tattatcccg	tattgacgcc	gggcaagagc	aactcggtcg	ccgcatacac	5760
	tattctcaga	atgacttggt	tgagtaactca	ccagtcacag	aaaagcatct	tacggatggc	5820
10	atgacagtaa	gagaattatg	cagtgtcc	ataaccatga	gtgataaacac	tgcggcaac	5880
	ttacttctga	caacatcggt	aggaccgaag	gagctaaccg	cttttttgc	caacatgggg	5940
	gatcatgtaa	ctcgccctga	tcttgggaa	ccggagctga	atgaagccat	accaaacgcac	6000
	gagcgtgaca	ccacatgcc	tgtacaaatg	gcaacaacgt	tgcgcaact	attaactggc	6060
	gaactacta	ctctacttcc	ccggcaacaa	ttaatagact	ggatgggggc	ggataaaatgt	6120
	gcaggaccac	ttctcgcttc	ggcccttccg	gttggctgg	ttatttgc	taaatctgg	6180
	gccgggtgagc	gtgggtctcg	cgttacatcc	gcagcactgg	ggccagatgg	taagcccttc	6240
	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggatgaacg	aaatagacag	6300
	atcgctgaga	tagtgcctc	actgattaaag	cattggtaac	tgtcagacca	agtttactca	6360
	tatatacttt	agattgattt	aaaacttcat	tttaattta	aaaggatcta	ggtgaagatc	6420
	ctttttgata	atctcatgac	aaaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	6480
	gaccggtag	aaaagatcaa	aggatcttct	ttagatcctt	tttttctgcg	cgtaatctgc	6540
	tgcttgcaaa	caaaaaaaacc	accgcttacca	gcgggtggttt	gttggccgga	tcaagagcta	6600
	ccaactcttt	ttccgaaggt	aactgggttc	agcagagcgc	agataccaaa	tactgtcctt	6660
	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgccc	tacatacaccc	6720
	gctctgtctaa	tcctgttacc	agtggctgt	gccagtgccg	ataagtctgt	tcttaccggg	6780
	ttggactcaa	gacgatagtt	accggataag	gcgcagcggt	cgggctgaaac	gggggggttc	6840
	tgcacacacg	ccagcttgg	gcgaacgacc	tacaccgaaac	tgagatcacct	acagcgttag	6900
	ctatgagaaa	gcgcacacgt	tcccgaaagg	agaaaaggccg	acaggtatcc	ggtaagcggc	6960
	agggtcgaaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	7020
	agtctctgtcg	gttttcgtcc	cctctgactt	gagcgtcgat	ttttgtgtat	ctcgctcagg	7080
	ggccggggcc	tatggaaaaa	cgccagaaac	ggccgtttt	tacggtctt	ggcccttttgc	7140
	tggcccttttgc	ctcacatgtt	ctttctctcg	ttatccccctg	attctgtgga	taaccgtatt	7200
	accgccttttgc	agttagctgt	taccgctcgc	cgcagccgaa	cgaccgagcg	cagcgagtc	7260
	gtgagcggagg	aagcggaaaga	gcccggaaata	cgcaaaccgc	ctctcccccgc	gcgttggccg	7320
	attcattaaat	gcagctggca	cgacagggtt	cccgaacttgc	aagcgggcag	tgagcgaac	7380
	gcaattaaatg	tgagtttagct	cactcattag	gcacccccagg	ctttacactt	tatgcttccg	7440
	gctcgatgtt	tgtgtggaaat	tgtgagcggtt	taacaatttc	acacaggaaa	cagctatgac	7500
	catgattaaatcg	ccaaatcgcc	gcgc				7523

40 <210> 80
<211> 4754
<212> DNA
<213> Artificial Sequence

45 <220>
<223> Description of Artificial Sequence: vector
pRK55pbsC31-Re

<400> 80	50	ggccggcccgatatgacacaa	gggggttggatccgtacacgg	ccgggggtggatccgtacacgg	cacgtacgcgatccgtacacgg	ggtgcttaacggatccgtacacgg	60
		accgtcagtcgccccggatccgtacacgg	gagaaattcgatccgtacacgg	gcccggatccgtacacgg	cacgtacgcgatccgtacacgg	cacgtacgcgatccgtacacgg	120
		ccaaacgaaatccgtacacgg	caaggccggatccgtacacgg	gacccgttccgtacacgg	ggaaatgtcgatccgtacacgg	ggccggatccgtacacgg	180
		gttgcgtcgatccgtacacgg	gcatttcaggatccgtacacgg	gaagcgccggatccgtacacgg	gcaacgtcgatccgtacacgg	gttcgggacgtccgtacacgg	240
	55	cgagatccgtacacgg	acgcattccgtacacgg	aacgaatgccatccgtacacgg	gcccggggcgatccgtacacgg	gctcaacatcgatccgtacacgg	300
		atgacgtgtcgcgttccgtacacgg	gcgttccgtacacgg	cgcctgaaggatccgtacacgg	tcatggacgcgttccgtacacgg	gattccgattgtctcgaaatccgtacacgg	360
		tgctcgccctggcgatccgtacacgg	ggcggtcgacgtccgtacacgg	attgtttccatccgtacacgg	ctcaggaaggatccgtacacgg	cgttcccgatccgtacacgg	420
		tcatggacgttccgtacacgg	gattcacccgtccgtacacgg	attatgcggcgtccgtacacgg	tgcacgcgtcgtccgtacacgg	gcacaaagaaatccgtacacgg	480
		agtccggcgaaatccgtacacgg	gattctcgacgtccgtacacgg	acgaagaaccgtccgtacacgg	ttcagcgcgtatccgtacacgg	attgggcgggtacgtccgtacacgg	540
	60	ggaaggccgttacgggtccgtacacgg	ttacgggttccgtacacgg	gagcttggatccgtacacgg	cgagacgaaatccgtacacgg	ggagatccacgtccgtacacgg	600
		gaatggtcaatgtcgttccgtacacgg	tgtcgttccgtacacgg	aacaaggttccgtacacgg	cgcaactcgatccgtacacgg	cactcccttaccggacacctccgtacacgg	660
		tccgagttcaatcccggacgtatccgtacacgg	gccccggacgtatccgtacacgg	atccgggttccgtacacgg	ggtggcgtgatccgtacacgg	gatcaagacgtccgtacacgg	720
		ttcccttcaatccggggcgtatccgtacacgg	gccggggcgtatccgtacacgg	caagccggcaatccgtacacgg	ttcacccgggtatccgtacacgg	cagcatcaggtccgtacacgg	780
		agcgcgtatggatccgtacacgg	cgctgacgtccgtacacgg	gtgccggacccatccgtacacgg	ggggcgagacatccgtacacgg	attgggaaagatccgtacacgg	840
	65	caaggccgttccgtacacgg	ggaccggatccgtacacgg	accgttatgtccgtacacgg	gaatccttccgtacacgg	ggaccggcgtatccgtacacgg	900
		tccgcccgttccgtacacgg	gggtgtatccgtacacgg	aagaagaagccgtacacgg	cgggacggcgtacacgg	ggccgaccacgtacacgg	960
		gttaccggatccgtacacgg	tcagcgcgacgtccgtacacgg	ccgatcaacgtccgtacacgg	tccggccggatccgtacacgg	cgagcttgcgtccgtacacgg	1020
		tcatcgagccatccgtacacgg	cgctgagttccgtacacgg	tatgagcttccgtacacgg	acgcgtatggatccgtacacgg	ggacggcgtatccgtacacgg	1080

5	aggggctttc	ccgggggcaa	gccattctgt	ccgcctatgg	caagctgtac	tgcgagtg	1140
	gcgcccgtcat	gacttcgaag	cgcggggaag	aatcgatcaa	ggactcttac	cgctccgtc	1200
	gccggaaagg	ggtcgaccgg	tccgcacctg	ggcagcacga	aggcacgtgc	aacgtcagca	1260
	tggccggact	cgacaagttc	gttgcgaaac	gcatcttcaa	caagatcagg	cacgccaaag	1320
	gcgacgaaga	gacgttggcg	cttctgtggg	aagccgcccc	acgcttcggc	aagctcaactg	1380
10	aggcgcctga	gaagagcggc	gaacggcga	accttggtc	ggagcgcg	gacgcccgt	1440
	acgcccctga	agagctgtac	gaagaccgg	cggcaggcgc	gtacgacgg	cccgttgca	1500
	gaaagcactt	ccggaaagcaa	caggcagcgc	tgaacgtccg	gcagcaaggg	gcggaagagc	1560
	gcttgcgca	acttgaagcc	gccgaagccc	cgaagcttcc	ccttgacc	tggttcccc	1620
	aagacgccc	cgctgaccgg	accggcccta	agtctgtgg	ggggcgcgcg	tcagtagacg	1680
15	acaagcgcgt	gttcgtcg	ctcttcgtag	acaagatcg	tgtaacgaa	tcgactacgg	1740
	gcagggggca	gggaacgccc	atcgagaagc	gcttcgtcgat	cacgtggcg	aagccgcga	1800
	ccgacgacg	cgaagacgac	gcccaggacg	gcacggaa	cgtacggcg	tagcggcgc	1860
	ccgggctoga	ggggggggcc	gttaccggc	ttttgttccc	tttagtgagg	gttaatttcg	1920
	agcttgggt	aatcatgttc	atagctttt	cctgtgtgaa	attgttatcc	gctcacaaat	1980
	ccacacaaca	tacgagccgg	aagcataaa	tgtaaagcc	gggggtcc	atagtgagc	2040
	taactcacat	taatttgcgt	gctcgtactg	cccgcttcc	agtcggggaa	cctgtcg	2100
	cagctgcatt	aatgaatcg	ccaaacgcg	ggggagggcg	tttgcgtat	ttggcgct	2160
20	tccgcgttcc	cgctcactg	ctcgctgc	tcggcgttcc	ggctgcgcg	agcggat	2220
	gctcactcaa	aggcggtaa	acggttatcc	acagaatcg	gggataacgc	aggaaagaac	2280
	atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcg	2340
	ttccataggc	tccgcccccc	tgacgagcat	cacaatcc	gacgtcaag	tcagagg	2400
	cgaaaaccga	caggactata	aagataccag	gctttcccc	ctggaa	cctcg	2460
25	tctccgttcc	cgaccctg	gcttacgg	tacctgtcc	ccttctcc	ttcgggaa	2520
	gtggcgctt	ctcatagctc	acgctgttag	tatctcgtt	cggtaggt	cgttcgt	2580
	aagctgggt	gtgtgcacga	acccccgtt	cagccccgacc	gctgcgc	atccgtaa	2640
	tatctgtt	agtccaaacc	ggtaagacac	gacttatcg	cactggc	agccact	2700
	aacaggattt	gcagagcgg	gtatgttagc	ggtgtacag	agttctt	gtgtggc	2760
30	aactacggc	acactagaag	gacagtattt	ggtatctgc	ctctgtgaa	gccagtt	2820
	tccggaaaaa	gagttggtag	ctcttgatcc	ggcaaaacaaa	ccaccgctt	tagcgg	2880
	ttttttgtt	gcaagcagca	gattacgc	agaaaaaaa	gatctcaag	agatcc	2940
	atottttcta	cggggctctga	cgctcagtgg	aacgaaaact	cacgttaagg	gattttgtc	3000
	atgagattt	caaaaaggat	cttcacccat	atcctttttaa	attaaaaatg	aagttttaaa	3060
35	tcaatctaa	gttatatga	gtaaacttgg	tctgacagt	accatgtt	aatcgtg	3120
	gcacctatct	cagcgatctg	tctatttcgt	tcatccatag	ttgcgtact	ccccgtcg	3180
	tagataacta	cgatacggg	gggttacca	tctggcccc	gtgcgtcaat	gataccgc	3240
	gaccacgct	caccggctcc	agatttata	gcaataaaacc	agccagccg	aaggccg	3300
	cgcagaagt	gtctgtcaac	titatccgc	tcatccagt	ctattaa	ttgcgggaa	3360
40	gctagagtaa	gtagttcg	agttatagt	ttgcgc	ttgttccat	tgctacagg	3420
	atcgtgggt	cacgctcg	gttttgtat	gcttcattc	gctccgg	ccaacgatc	3480
	aggcgagtt	catgatcccc	catgttgtc	aaaaaagcg	ttagctc	cggtc	3540
	atcgttg	gaagtaagtt	ggccgc	ttatcact	tggttatgg	agcactgt	3600
	aattcttta	ctgtcatgc	atccgtaa	tgctttct	tgactgg	gtactca	3660
45	aagtcttct	gagaatagt	tatgcggc	ccaggttgc	cttgc	gtcaata	3720
	gataatacc	cgccacatag	cagaactt	aaagtgc	tcattgg	acgttct	3780
	gggcgaaaaac	tctcaaggat	cttaccgt	ttgagatcc	gttcgat	accactgt	3840
	gcacccaa	gatctcagc	atctttact	ttcaccag	tttctgg	agcaaaa	3900
	gaaaggcaaa	atggccaaa	aaagggaata	agggc	gaaatgtt	aataact	3960
50	ctcttcctt	ttcaatatta	ttgaagcatt	tatcagg	attgtct	gagcgg	4020
	atatttgaat	gtatgtt	aaataaaca	atagggg	ttt	aaaa	4080
	gtgccaccta	aattgt	gttaatattt	tgtt	cgcgtaa	ccccgaaaa	4140
	tcagcttatt	tttt	taggccc	aaa	tcc	ttttgtt	4200
	agaccggat	accgg	gttgc	aaaa	tataa	aaaaa	4260
55	tggacttccaa	cgtcaaaagg	cgaaaaacc	tctat	cgatgg	ccatcg	4320
	catcacccat	atcaagttt	ttgggtt	cggt	ggcc	atcg	4380
	aaggggcccc	ccgattt	gttgc	gg	actaa	ccggaa	4440
	gaaagaaagc	gaaaggag	gcttgc	gg	aaacgt	ccgg	4500
	taaccaccac	acccgc	cttaat	tcgt	gg	ccgtc	4560
60	ggctgcgca	ctgttgg	gggcgat	tcgtt	gg	ccaggt	4620
	cgaaaagg	atgtgt	aggcgat	ttcg	ttt	ccagtc	4680
	gacgtt	aa	atcgact	atc	atagg	gttgc	4740
	tccaccgc	cc	atgtgtt	ttt	ttt	ccatcg	4754

65 <210> 81
<211> 4773
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: vector
5 pRK63pbsC31NLS-Re

<400> 81

gcccccacca tgcccaagaa gaagaggaag gtgacacaag gggttgtgac cggggtggac 60
 acgtacgcgg gtgcttacga ccgtcagtcg cgcgagcgcg agaattcag cgcagcaagc 120
 10 ccagcgacac agcgtagcgc caacgaagac aaggcggccg accttcagcg cgaagtccgag 180
 cgcgacgggg gccgggtcag gttcgtccgg catttcagcg aagcgcggg cactgtccgg 240
 ttccggacgg cggagcgcgg ggagttcga cgcattctga acgaatccg cgcggggccgg 300
 ctcaacatga tcattgtcta tgacgtgtcg cgcttctcg gcctgaaggt catggacgcg 360
 attcggattt tctcggattt gtcgcctg ggcgtacga ttgtttccac tcaggaaggc 420
 15 gtcgttccggc agggaaacgt catggacgtt attcacctga ttatgcgggtt cgacggtcg 480
 cacaagaat ttgcgttcaa gtcggcgaag atttcgaca cgaagaacct tcagcgcgaa 540
 ttgggcgggtt acgtcggcgg gaaggccct tacggcttc agcttgcgtt ggagacgaag 600
 gagatcacgc gcaacggccg aatggtcaat gtcgtcatca acaagcttc gcaactcgacc 660
 20 atcaagacgc acaaaccactt tcccttcaag cggggcagtc aagccgcatt tcacccgggc 780
 agcatcacgg ggcttgcattt ggcgttgcg gtcgtacgcgg tgccgacccgg gggcgagacg 840
 attgggaaga agaccgcctt aagcgcctgg gaccggcaaa cgcgtatgcg aatcttcgg 900
 gaccggcgtt ttgcgggcattt cgccgttgcg gtgtatctaca agaagaagcc ggacggcacc 960
 25 cogaccacga agattgaggg ttaccgcattt cagcgcgacc catgcggccg gtcgtatgtt atgagcttca ggcgtgggtt 1080
 gagcttgcattt gcggttgcattt catcgaggcc gtcgtatgtt ccatttcgtt cgcgtatggac 1140
 gacggcgggg ggcgcggcaaa ggggtttcc cggggggcaag aagctgtactt gcggttgcattt gtcgtatgtt atcgatcaag 1200
 30 gactcttacc gtcggcgtcg gccggaaatgtt gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgtt gtcgtatgtt 1260
 ggcacgtca acgtcgtatcc ggcggcactt gacaaatgtt ccgcgtatgcg gtcgtatgtt 1320
 aagatcacggc acgcggcaagg ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1380
 35 cgcttcggca agctcacttgc ggcggcgttcc gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1440
 gagcggccgg gaccccttgc gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1500
 tacgacggac ccgttggcag gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1560
 cagcaagggg cgaaagagcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1620
 40 cttgaccaat gtttcccgaa agacggccgac gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1680
 gggcgcgcgtt cagtagacga caagcgcgtt gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1740
 gtcacgaagt cgactacggg cagggggcgtt gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1800
 45 acgtgggcga agccggcgcac gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1860
 gtagcggcgtt aggccggcgc ttcgtatgtt ccgcgtatgcg gtcgtatgtt 1920
 50 ttagtgagggtt ttaatttgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 1980
 ttgttatccg ctcacaatttccg cacaacat acgacggcgtt gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2040
 ggggtcctaa tgagtgagct aactcacattt aatttcgttccg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2100
 gtcgggaaac ctgtcgttgcg agctgcattt atgaaatcgcc gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2160
 55 tttgcgtattt gggcgcttccg ccgttccctc gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2220
 gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2280
 ggataacgcgca ggaaagaaacca gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2340
 60 ggcgcgttccg ctggcggtttt acgctcaagt cagaggtggc gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2400
 tggaaatgttccg ctcgtgcgtt gaaacccgac aggtactataa gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2460
 65 ctttctccct tcggaaagcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2520
 ggtgttaggtc gttcgcttca tccataggtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2580
 ctgcgccttca tccgttactt acaggattttt ctcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2640
 actggcagca gccactggta gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2700
 gttcttgaag tgggtggctt tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2760
 70 tctgttgcgtt ccgttacccat tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2820
 caccgttgcgtt agcgttgcgtt tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2880
 atctcaagaa gatccttgcg tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 2940
 acgtttaagggtt attttgcgttccg tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3000
 75 taaaaatgttccg tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3060
 ccaatgttccg tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3120
 tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3180
 tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3240
 tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3300
 80 tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3360
 tattaaattgtt ccgttgcgttccg tccgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3420
 85 tgggttccgattt gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3480
 ctccgttccgatccaa caacatgttccg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3540
 tagctccatccgatccaa gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt ccgcgtatgcg gtcgtatgtt 3600

5 ggttatggca gcactgcata attcttacat tgcattgtccaa tccgttaagat gctttctgt 3660
 gactggtag tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc 3720
 ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa aagtgcctat 3780
 cattggaaaaa cgttcttcgg ggcgaaaact ctcaggatc ttaccgtgt tgagatccag 3840
 ttgcgtgtaa cccactcgta caccctactg atcttcagca tctttactt tcaccagcg 3900
 ttctgggtga gcaaaaacag gaaggcaaaaa tgccgcaaaa aagggaaataa gggcgacacg 3960
 gaaatgttga atactcatac tcttccttt tcaatattat tgaagcattt atcagggtta 4020
 ttgtctcatg agcggataca tatttgaatg tatttagataa aataaacaataa taggggttcc 4080
 10 ggcgcacattt cccggaaaaa tgccacctaa atttgaagcg ttaatatttt gttaaaaattt 4140
 gctttaaattt tttgttaaat cagtcattt ttttaccaat aggccgaaaat cggcaaaaatc 4200
 ctttataat caaaagaata gaccgagata ggggtgagtg ttgttcoagt ttggaaacaag 4260
 agtccactat taaagaacgt ggactccaaac gtcaaagggc gaaaaacgt ctatcagggc 4320
 gatggcccac tacgtgaacc atcaccctaa tcaagtttt tggggtcgag gtgcgttaaa 4380
 15 gcaactaaatc ggaaccctaa agggagcccc cgatttagag ctgtacgggg aaagccggcg 4440
 aacgtggcga gaaaggaagg gaagaaagcg aaaggagcgg ggcgttagggc gctggcaagt 4500
 gtagcggtca cgctgcgtt aaccaccaca cccgcggcgc ttaatgcgcc gtcacagggc 4560
 gctgtccattt cgccatttcag gctgcgcac tggggaaag ggcgatcggt gggggctct 4620
 tcgctattac gccagctggc gaaagggggta tggctgtcaa ggcgattaag ttgggttaacg 4680
 20 ccagggtttt cccagtcacg acgttgtaaa acgacggcca gtgaattgtt atacgactca 4740
 ctataggcg aattggagct ccaccgggtt ggc 4773

25 <210> 82
 <211> 7803
 <212> DNA
 <213> Artificial Sequence

 <220>
 30 <223> Description of Artificial Sequence: vector
 pPGKattA1

 <400> 82
 tatacatgtct ggatccgcgt taacacctaa gaaggcgaag ttttccttac accttgcaga 60
 tataaagtgt ctaacagttt aaaatatccg atgaggcata tttatgttgg acccgtagct 120
 35 cagccaggat agagcactgg cctccggagc cggaggtccc ggttcaaat cccggcgggt 180
 ccgttatatta ctttttgcattt cagatttagat ttgttaatct ttattacaag gataatttga 240
 tcttgcattat tggtaactctt ctactctata atttttatga gaaattcaca gtcgtccctt 300
 tataccataaa atagctaagt ttgtcaaaatg tcttattaaa ctctccatgt agagattaaa 360
 40 tcggatccat aacttcgtat agcatacatt atacgaaatg ataccgggcc accatggtcg 420
 cgagtagctt ggcactggcc gtcgttttac aacgtcgtaa ctggggaaaac cctggcgtaa 480
 cccaaacttacatc tgccttgcgca gcacatcccc ctttcggccat ctggcgtaat agcgaagagg 540
 cccgcacccgatc tgccttgcgca caacagttgc gcagcctgaa tggcgaatgg cgctttgcct 600
 ggttccggcc accagaagcg gtggccggaaa gctggctgga gtgcgatctt cctggggccg 660
 atactgtcgatcgtccctca aactggcaga tgcacggta cgtgcgccc atctacacca 720
 45 acgttaaccta tcccttacatcgtcaatccgc cgtttgttcc caccggaaat ccgcacgggtt 780
 gttactcgct cacatttaat gttgatgaaa gctggctaca ggaaggccag acgcgaattt 840
 tttttgatgg cgttaactcg gcttttcatc tgggtgtcaa cgggcgttgc gtcgttacg 900
 gccaggacacatc tgccttgcgca tctgaatttgc acctgagcgtt attttacgc gccggagaaa 960
 50 accgcctcgc ggtgatggt ctgcgttggaa gtgacggcag ttatctggaa gatcaggata 1020
 tggccggat gacggcattt tccgtgacg tctcggttgc gctaaaaccg actacacaaa 1080
 tcagcgattt ccatgttgcgactcgctttaat atgatgattt cagccgcgt gtactggagg 1140
 ctgaagttca gatgtgcggc gagttgcgtg actacatcgtt ggttacatg tttttatggc 1200
 aggggtgaaac gcaggcgcc acggcaccgc cgccttcgg cggtaattt atcgatgagc 1260
 55 gtgggtgttgc tggcgatcgc gtcacactac gtcgttgcgtt cggaaaccgcg aaactgtgga 1320
 ggcgcgaaat cccgaatctc ttcgtgcgg tgggtgttgc gacacccgcg gacggcacgc 1380
 tgattgttgcgactcgatcgc gatgtcggtt tccgcggatgtt gggattgaa aatggctgc 1440
 tgcgtgttgcgactcgatcgc ttcgtgttgc gggcgatgtt ccttcgcgcg catcatcctc 1500
 tgcgtgttgcgactcgatcgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1560
 60 acaactttaa cggccgtgcgca ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1620
 ggcgcacccgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1680
 caatgaatcg ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1740
 gaatgggttgcgactcgatcgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1800
 caggccacccgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1860
 gcccgggttgcgactcgatcgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1920
 65 tgcgtgttgcgactcgatcgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 1980
 aatggctttc gtcaccttgcgactcgatcgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 2040
 tgggttaacatcgttgcgactcgatcgc ttcgtgttgc gggcgatgtt ctttcgcgcg catcatcctc 2100

	tacagggcgg	cttcgctcgg	gactgggtgg	atcagtgcgt	gattaaatat	gataaaaacg	2160
	gcaaccctgt	gtcggtttac	ggcggtgatt	ttggcgatac	gccgaacgt	cggcagttct	2220
	gtatgaacgg	tctggcttt	gccgaccgca	cgccgcattcc	agcgctgacg	gaagcaaaac	2280
5	accagcaga	gttttccag	ttccgtttat	ccgggcaaaac	catcgaaatg	accagcgaat	2340
	acctgttccg	tcatagcgat	aacgagctcc	tgcactggat	ggtggcgctg	atggtaagc	2400
	cgctggcaag	cggtgaagt	cctctggatg	tcgctccaca	aggtaaacag	ttgattgaac	2460
	tgccctgaact	accgagccg	gagagccg	ggaactctg	gctcacagta	cgcgtatgtc	2520
10	aaccgaacgc	gaccgtatgg	tcagaagccg	ggcacatcg	cgcctggcag	cagtggcgtc	2580
	ttggcggaaa	cctcagtgt	acgctccccc	ccgcgtccca	cgcacatccc	catctgacca	2640
	ccagcgaat	ggattttgc	atcgagctgg	gtaataaagcg	ttggcaattt	aaccgcagg	2700
	caggctttct	ttcacagat	ttggatggcg	ataaaaaaaca	actgctgacg	cgcgtgcgcg	2760
	atcagttcac	ccgtgcaccc	ctggataaag	acattggcg	aagtgaagcg	acccgcattt	2820
15	accctaacgc	otgggtcgaa	cgctggaaagg	cggggggcca	ttaccaggcc	gaagcagcgt	2880
	tgttcagtg	cacggcagat	acacttgcgt	atgcggtgct	gattacgacc	gctcacgcgt	2940
	ggcagcatca	ggggaaaacc	ttatattatca	gcccggaaaac	ctacggatt	atggtagtg	3000
	gtcaaattgc	gattaccgtt	gtgttgaag	ttggcgagcg	tacaccgc	cggcgcgg	3060
	ttggcctgaa	ctgcagctg	gcccaggtag	cagagcgggt	aaactggctc	ggattagggc	3120
	cgcaagaaaa	ctatccgcac	cgccctactg	ccgcctgttt	tgaccgtgg	atctgccc	3180
20	tgtcagacat	gtataccccg	tacgcttcc	cgagcgaaaa	cggtctgcgc	tgcgggacgc	3240
	gcaatttggaa	ttatggccca	caccagtggc	gcggcgactt	ccagttcaac	atcagccgt	3300
	acagtcaaca	gcaactgtat	gaaaccagcc	atcgccatct	gctgcacgc	gaagaaggca	3360
	catggctgaa	tatcgacgg	ttccatatgg	ggattggcg	cgacgactcc	ttggagccgt	3420
	cagtatcggc	ggaattccag	ctgagcccg	gtcgctacca	ttaccagtt	gtctgggtgc	3480
25	aaaaataata	ataaccgggc	agggggatc	tttgtgaagg	aaccttactt	ctgtgggtgt	3540
	acataatttg	acaaactacc	taacagagatt	taaagctcta	aggtaaatat	aaaattttta	3600
	agtgtataat	gtgttaaact	actgattcta	atgtttgtg	tattttagat	tccaaacctat	3660
	gaaactgtat	aatgggagca	gtggtggaaat	gcoagatcca	gacatgataa	gatacatatga	3720
	tgagtttga	caaaccacaa	ctagaatgca	gtaaaaaaaaa	tgcttattt	gtgaaatttg	3780
30	tgatgttatt	gctttatgtt	taaccattat	aagctgcaat	aaacaagtt	acaacaacaa	3840
	ttgcattcat	tttatgtttc	aggttcagg	ggaggtgtgg	gaggttttt	aaagcaagta	3900
	aaacctctac	aaatgtggta	tggctgatta	tgatctgcgg	ccgcaggggcc	tcgtgataacg	3960
	cctattttt	tagttaatg	tcatgataat	aatggtttct	tagacgtcg	gtggcactt	4020
	tcggggaaat	gtgcgcggaa	cccctattt	tttattttt	taatacatt	caaatatgt	4080
35	tccgctcatg	agacaataac	cctgataat	gcttcaataa	tattgaaaaa	ggaagagtat	4140
	gagtattcaa	cattccgtg	tcgccttat	tcctttttt	gcggcatttt	gccttcctgt	4200
	tttgcgtcac	ccagaaacgc	tggtggaaat	aaaagatgct	gaagatcagt	tgggtgcacg	4260
	agtgggttac	atcgaaactgg	atctcaacag	cggtaaagatc	cttgagagtt	ttcgccccga	4320
	agaacgttt	ccaatgtat	gcaactttaa	agttctgtc	tgtggcgcgg	tattatcccg	4380
40	tattgacgcc	ggcaagagac	aactcggtcg	cgcatacac	tattctcaga	atgacttgg	4440
	tgagttactca	ccagtcacag	aaaagatct	tacggatggc	atgacagttaa	gagaattatgt	4500
	cagtgtgtcc	ataaccatgt	gtgataacac	tgccgccaac	ttacttctga	caacgatcgg	4560
	aggaccgaa	gagctaaccg	cttttttgc	caacatgggg	gatcatgtaa	ctcgccttga	4620
	tcgttgggaa	ccggagctga	atgaacccat	accaaaccgc	gagcgtgaca	ccacgatgcc	4680
45	tgttagcaat	gcaacaacgt	tgcgcaaaact	attaactggc	gaactactta	ctctagtttc	4740
	ccggcaacaa	ttaatagact	ggatggaggc	ggataaaagt	gcaggaccac	ttctgcgtc	4800
	gcccttcgg	gctggctgg	ttattgtga	taaattctgga	gccggtgagc	gtgggtctcg	4860
	cggtatcatt	gcagcactgg	ggccagatgg	taagccctcc	cgtatctgt	ttatctacac	4920
	gacggggagt	cagggcaacta	tggatgaacg	aaatagacag	atcgctgaga	taggtgcctc	4980
50	actgattaa	cattggtaac	tgtcagacca	agtttactca	tatatacttt	agattgatt	5040
	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	cttttgata	atctcatgac	5100
	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	gacccctgt	aaaagatcaa	5160
	aggatcttct	tgagatcctt	tttttctcg	cgtaatctgc	tgcttgcaaa	caaaaaaaaaacc	5220
	accgctacca	gccccgttt	gtitgcccga	tcaagagcta	ccaaactctt	ttccgaaggt	5280
55	aactggcttc	agcaagcgc	agataccaaa	tactgtcctt	ctagtgtac	cgtagttagg	5340
	ccaccacttc	aagaactctg	tagcaccgc	tacatacctc	gctctgctaa	tcctgttacc	5400
	agtggctgt	gccagtggcg	ataagtctgt	tcttaccggg	ttggactcaa	gacgatagtt	5460
	accggataag	gcgcagcggt	cgggctgaaac	gggggggttcg	tgcacacagc	ccagcttgc	5520
	gcgaacgacc	tacaccgaa	ttagatacttcc	acagcgtgag	ctatgagaaa	gcgcacacgt	5580
60	tcccgaaggg	agaaaaggcgg	acaggtatcc	ggttaagcggc	agggtcggaa	caggagagcg	5640
	cacgagggg	cttcaggggg	gaaacgcctg	gtatctttat	atgtctgtcg	gtttcgc	5700
	cctctgtactt	gagcgctcgat	ttttgtgtat	ctcgatcagg	ggggcggagcc	tatggaaaaaa	5760
	cgccagcaac	gccccctttt	taacgttcc	ggccttttgc	tggccttttgc	ctcacatgtt	5820
	ctttcctgcg	ttatcccctg	attctgttgc	taaccgtatt	accgccttttgc	atgtgagctga	5880
	taccgtctgc	cgcaaggccaa	cgaccgagcg	cagcgagtc	gtgagccagg	aagcggaaaga	5940
65	gcccccaata	cgcaaccgc	ctctccccgc	gcttgcggcc	attcattaat	gcagctggca	6000
	cgacagggtt	cccgactgg	aagcgggcag	ttagcgcac	gcaattaatg	ttagtttagt	6060
	cactcattag	gcacccagg	ctttacactt	tatgttccg	gtctgtatgt	tgtgtggaaat	6120

5 tggtagcgga taacaatttc acacaggaaa cagctatgac catgattacg ccaagctggc 6180
 gcttcgagga attctaccgg gtaggggagg cgctttccc aaggcagtct ggagcatgc 6240
 ctttagcagc cccgcgtggc acttggcgt acacaagtgg cctctggctc gcacacattc 6300
 cacatccacc gtaggcgcc aaccggctc gtttttggt ggcggctcg cgccaccc 6360
 tactccccc ttagtcagga agttcccccc cgccccgcag ctgcgtcgt gcaggacgtg 6420
 acaaattggaa gtagcacgtc tcactagtct cgtcagatg gacagcaccc ctgagcaatg 6480
 gaagcgggta ggccttggg gcagcggca atagcagct tgctccctcg ctttctggc 6540
 tcagaggctg ggaaggggtg ggtccgggg cggctcagg ggcggctca ggggcccc 6600
 10 ggcggcccgaa aggtccctcg gaggccccc attctgcacg cttcaaaagc gcacgtctc 6660
 cgcgctgttc tcctttctc catctccccc ctttgcacc tgcagcccg tacagttcg 6720
 ataacttcgt atagcataca ttatacgaag ttataagett gcatgcgtc aggtcggccg 6780
 ccacgacccg cccggccgtg ccgcacccat cccctgaccc acgcccctga cccctcacaa 6840
 ggagacgacc tttcatgacc gatcacaacg ccacggcgtcg cctcgccacc cgcgacgacg 6900
 15 tccccccggc cgtacgcacc ctgcggccg cgttcgccg ctacccggc acgcgcacca 6960
 cctgcgaccc ggaccgcac atcgagcggt tcaccgagct gcaagaactc ttccctacgc 7020
 gcttcgggct cgacatcgcc aagggtgtgg tcggcgacga cggcgcccg gtggcggtct 7080
 ggaccacgccc ggagacgctc gaagcggggg cgggttgcgc ctagatcgcc cgcgcgtatgg 7140
 20 ccgagtttag cggttcccg cttggccgcg agcaacagat ggaaggcctc ctggcgcccg 7200
 accggcccaa ggagcccggt tggttctgg ccaccgtcg cgtctcgccc qaccaccagg 7260
 25 gcaagggtct gggcagcggc gtcgtgtcc ccggagtgga ggcggcccgag cgccgcgggg 7320
 tgcccgccctt cctggagacc tccgcgcacc gcaacctccc ctctacgc cggctcggtct 7380
 tcaccgtcac cgccgacgtc gagtccccga aggaccgcgc gacctgggtc atgaccgc 7440
 agcccggtgc ctgacgcacc cccacgcacc cgcagcgcgg acccgaaagg agegcacgac 7500
 30 cccatggctc cgaccgaagc cgaccggc ggcccgcccg accccgcacc cgccccccgag 7560
 gcccaccgac tctagaggat cataatcagc cataccacat ttgttagagggtttacttgc 7620
 ttaaaaaacc tcccacacct cccctgtAAC ctgaaacata aaatgaatgc aattgttgc 7680
 gtttaacttgt ttattgcagc ttataatggt tacaataaa gcaatagcat cacaatatttc 7740
 acaaataaaag catttttttc actgcattct agttgtggtt tgtccaaact catcaatgt 7800
 tct 7803

35 <210> 83
 <211> 8167
 <212> DNA
 <213> Artificial Sequence

40 <220>
 <223> Description of Artificial Sequence: vector
 pPGKattA2

45 <400> 83
 tatacatgtct ggatccgcgt taacacctaa gaaggcgaag ttttccttac accttgcaga 60
 tataaaagtgt ctaacagttt aaaatatccg atgaggcata tttatgttgg acccgtagct 120
 cagccaggat agagcactgg cctccggcgc cggaggctcc gggttcaaat cccggcggt 180
 50 cgttatattt ctttttggatt gagatttagat ttgttaatct ttatataag gataatttga 240
 tottgtatattt tggttaactct ctacttata atttttatga gaaatttcaca gtcgtccctt 300
 tataccataa atagctaagt ttgtcaaaatg tcttattaaa ctctccatgt agagattaaa 360
 tcggatccat aacttcgtat agcatacatt atacgaagtt ataccggcc accatggcgt 420
 55 cgagtagctt ggcactggcc gtcgttttac aacgtcgtga ctggggaaaac cctggcggt 480
 cccaacttaa tcgccttgca gcacatcccc ctttcggccag ctggcgtaat agcgaagagg 540
 cccgcaccgaa tcgccttcc caacagtgc gcagcctgaa tggcgaaatgg cgctttgcct 600
 gtttccggc accagaagcg gtggcgaaa gctggctggaa gtgcgtatctt cctgaggccg 660
 atactgtcgt cgtccctca aactggcaga tgcacggta cgatgcgc accatggcgt 720
 60 acgtaaccta tcccattacg gtcaatccgc cgtttgttcc caccggaaat ccgacgggtt 780
 gttactcgct cacatttaat gttgatggaa gctggctaca ggaaggccag acgcgaat 840
 ttttgatgg cgttaactcg gcgtttcatc tgggtgcacg cgggcgttgg gtgggttacg 900
 gccaggacacg tcgtttggcc tctgaatttgc acctgagcgc atttttacgc gccggagaaa 960
 accgcctcgc ggtgatgggt ctgggttggaa gtgcacggcag ttatctggaa gatcaggata 1020
 tggcggttgc gaggccattt ttccgtgacg tctcgttgcgt gcataaaccg actacacaaa 1080
 65 tcagcgattt ccatgttgcgc actcgctta atgatgattt cagccgcgt gtactggagg 1140
 ctgaagttca gatgtgcggc gagttgcgtg actacactacg ggttaacagtt tctttatggc 1200
 agggtaaaac gcaggatcgcc acggccaccc cgcccttccgg cggtaaattt atcgatgacg 1260
 gttgtgggttgc tgccgatcgcc gtccacactac gtcgtacatcg cggaaaaccg aactgtgg 1320
 gcccggaaat cccgaatctc tatacggttgcgg tgggtgaact gcacaccggc gacggcacgc 1380
 tgattgaagg agaaggctgc gatgtcggtt tccgcgaggt gcggttggaa aatggtctgc 1440
 tgcgtctgaa cggcaagccg ttgtgtatc gagcagacga tgggtcagga cctgcacgag catcatcc 1500
 tgcattgtca ggtcatggat gagcagacga tgggtcagga tatacctgctg atgaaggcaga 1560

5	acaactttaa	cgccgtgcgc	tgttcgcatt	atccgaacca	tccgctgtgg	tacacgcgt	1620
	gcgaccgcta	cggcctgtat	gtggtgatg	aaggccatat	tgaaacccac	ggcatggtc	1680
	caatgaatcg	tctgaccgat	gatccgcgct	ggctaccggc	gatgagcgaa	cgcgtaacgc	1740
	gaatggtgca	gcgegatcgt	aatcaccgg	gtgtgatcat	ctggctgctg	gggaatgaat	1800
	caggccacgg	cgctaattcac	gacgcgtgt	atcgctggat	caaatctgtc	gatcccttccc	1860
	gcccggtgca	gtatgaaggc	ggcggagccg	acaccacggc	caccgatatt	atttgcggaa	1920
	tgtacgcgct	cgtggatgaa	gaccagccct	tcccggctgt	gccgaaaatgg	tccatcaaaa	1980
	aatggcttc	gctacctgga	gagacgcgccc	cgctgatcct	ttgcgaatac	gcccacgcga	2040
10	tggttaacag	tcttggcggt	ttcgctaaat	actggcaggc	gtttcgtcag	tatccccgtt	2100
	tacagggcgg	cttcgtctgg	gactgggtgg	atcagtcgt	gattaaatat	gatgaaaacg	2160
	gcaaccctgt	gtccgcttac	ggcgggtatt	ttggcgatac	gcogaacgat	cgccagttct	2220
	gtatgaacgg	tctgggttctt	gcccgcacca	cgccgcaccc	agogctgacg	gaagcaaaac	2280
	accagcagca	gttttccag	ttccgtttat	ccgggcaaac	catcgaaatg	accagcgaat	2340
	acctgttccg	tcatacgat	aacgagctcc	tcgactggat	ggtggcgctg	gatggtaagc	2400
	cgtgtccaaag	cggtaagtg	cctctggatg	tgctccac	aggttaacaaag	ttgatttaac	2460
	tgcctgaact	accgcacccg	gagagcgcgg	ggcaactctg	gctcacagta	cgcgtatgtc	2520
15	aacgcacccg	gaccgcatgg	tcagaagccg	ggcacatcag	cgccctggcag	cagtggcgctc	2580
	ttggcggaaaaa	cctcagtgtg	acgtccccc	cogcgtccca	cgccatcccc	catctgacca	2640
	ccagcgaaaat	ggattttgc	atcgagctgg	gtataaagcg	ttggcaattt	aaccgcagg	2700
	caggcttct	ttcacagatg	tggattggcg	ataaaaaaaca	actgctgacg	ccgctgcgcg	2760
20	atcagttcac	ccgtgcacccg	ctggataaac	acattggcgt	aagtgaagcg	accgcatttg	2820
	accctaaccgc	ctgggtcgaa	cgctggagg	cgcgccccca	ttaccaggcc	gaagcagcgt	2880
	tgttgcaagt	cacggcagat	acacttgctg	atcggtgtc	gattacgacc	gtcacaagg	2940
	ggcagcatca	ggggaaaacc	ttatattatca	gccggaaaac	ctaccggatt	gatggtagtg	3000
25	gtcaaatggc	gattaccgtt	gatgttgaag	tggcgagcga	tacaccgcat	ccggcgccgaa	3060
	ttggcctgaa	ctgcccagctg	gcccaggtag	cagagcgggt	aaactggctc	ggatttagggc	3120
	cgcaagaaaaa	ctatcccgac	cgccttactg	ccgcctgttt	tgaccgctgg	gatctgccat	3180
	tgtcagacat	gtataccccg	tacgittcc	cgagcggaaa	cggtctgcgc	tgccggacgc	3240
30	gcaattgaa	ttatggccca	caccagtggc	gccccgactt	ccagttcaac	atcagccgt	3300
	acagtcacaa	gcaactgtat	gaaaccagcc	atcgccatct	gctgcacgcg	gaagaaggca	3360
	catggctgaa	tatcgacgg	ttccatatgg	ggatgtgtgg	cgacgactcc	tggagcccg	3420
	cagtatccgc	ggaattccag	ctgagcgcgc	gtcgtacca	ttaccagttg	gtctgggtgc	3480
35	aaaaataata	ataaccgggc	aggggggatc	tttgcgtgg	aaccttactt	ctgtgggtgt	3540
	acataatttg	acaaactacc	taacagatgg	taaagctcta	agttaaatat	aaaattttta	3600
	agtgtataat	gtgttaaact	actgattcta	attgtttgt	tattttagat	tccaaacctat	3660
	gaaactgtat	aatggggaca	gtgtgtggat	gcccagatcca	gacatgataa	gatacatgta	3720
40	tgagtttgg	caaaccacaa	ctagaatgc	gtgaaaaaaaaa	tgcttttattt	gtgaaatttg	3780
	tgatgtctt	gttttatttg	taaccattat	aagctgcaat	aaacaaggta	acaacaacaa	3840
	ttgcatttc	tttatgtttc	aggttcaggg	ggaggtgtgg	gaggtttttt	aaagcaagta	3900
	aaacctctac	aatgtggta	tggctgatta	tgtatctgcgg	ccgcaggggcc	tcgtgatacg	3960
45	cotattttt	taggttaatg	tcatgataat	aatggtttct	tagacgtcag	gtggcactt	4020
	tcggggaaat	gtgcgoggaa	cccctatttg	tttatttttc	taaatacatt	caaataatgta	4080
	tccgctcatg	agacaataac	cctgataaaat	gcttcaataa	tattgaaaaa	ggaagagtat	4140
	gagattccaa	cattccgtg	tcgcccatt	tccctttttt	gcccattttt	gccttcctgt	4200
50	ttttgctcac	ccagaaacgc	tggtgaaagt	aaaagatgt	gaagatcagt	tgggtgcacg	4260
	agtgggttac	atcgaaactgg	atctcaacag	cggtaaagatc	cttggagagtt	ttagccccga	4320
	agaacgtttt	ccaaatgtat	gcactttaa	agttctgcta	tgtggcgcgg	tattatcccg	4380
	tattgacgaa	ggcaagagc	aactcgtcg	ccgcatacac	tatttcaga	atgacttgg	4440
55	tgagtactca	ccagtcacag	aaaagcatct	tacggtatggc	atgacagtaa	gagaattatg	4500
	cagtgtgcc	ataaccatga	gtgataacac	tgccggccaa	ttacttctga	caacgatccg	4560
	aggaccgaag	gagctaaccg	cttttttgc	caacatgggg	gatcatgtaa	ctcgccctg	4620
	tcgttgggaa	ccggagctga	atgaaggccat	accaaaccgc	gagcgtgaca	ccacgatgcc	4680
	tgttagcaatg	gcaacaacgt	tgccgaaact	attaacttgc	gaactactta	ctctagtttc	4740
60	ccggcaacaa	ttaatagact	ggatggaggc	ggataaaagg	gcaggaccac	ttctgcgttc	4800
	ggcccttccg	gctggctgg	ttatgtctga	taaatctgga	gcccgtgagc	gtgggtctcg	4860
	cggatcatt	gcagacttgc	ggccagatgg	taagccctcc	cgtatctgt	ttatctacac	4920
	gacggggagt	caggcaacta	tggatgtac	aaatagacag	atcgtgtgaa	taggtgcctc	4980
	actgattaag	catgtttaac	tgtcagacca	agtttactca	tataactt	agattgattt	5040
65	aaaacttcat	tttaatttta	aaaggatcta	ggtgaagatc	tttttgata	atctcatgac	5100
	caaaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	gaccggtag	aaaagatcaa	5160
	aggatcttct	tgagatcctt	tttttctgcg	cgtaatctgc	tgcttgcata	caaaaaaacc	5220
	accgctacca	gcccgtgttt	gtttgcgg	tcaagagct	ccaaactctt	ttccgaaggt	5280
	aactggcttc	agcagagcgc	agataccaa	tactgtcctt	ctagtgttag	cgtagttagg	5340
	ccaccacttc	aagaactctg	tagcacggc	tacataccct	gctctgtcaa	tcctgttacc	5400
	agtggctgt	gccagtgccg	ataagtctg	tcttacccgg	ttggactcaa	gacgatagtt	5460
	accggataag	gcccgcgggt	cgggctgaac	gggggggttcg	tgcacacagc	ccagcttgg	5520
	gccaacgcacc	tacaccgaac	tgagataacct	acagcgttag	ctatgagaaaa	gcgccacgc	5580

<210> 84

<210> 51

<212> DNA

50 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer XisA1

55 <400> 84

ataagaatgc ggcgcggcga tatgcaaaat cagggtcaag acaaataatca a

51

<210> 85

<211> 76

<212> DNA

<213> Artificial Sequence

65 <220>

<223> Description of Artificial Sequence: primer XisA2

<400> 85

80

ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgcaaaatc agggtaaga 60
caaatatcaa caagcc 76

5 <210> 86
<211> 44
<212> DNA
<213> Artificial Sequence

10 <220>
<223> Description of Artificial Sequence: primer XisA3

<400> 86
ataagaatgc ggccgctcaa ctattcttat aagctatttc catc 44

15 <210> 87
<211> 82
<212> DNA
20 <213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer nifD1

25 <400> 87
cgatggctct tcccttccgt caaatgcact cttgggatta ctccgaacct agcgatgggg 60
tgcaaatgtc agatcagata ag 82

30 <210> 88
<211> 82
<212> DNA
<213> Artificial Sequence

35 <220>
<223> Description of Artificial Sequence: primer nifD2

<400> 88
cgcttatctg atctgacatt tgcacccat cgcttagttc ggagtaatcc caagagtgca 60
tttgacggaa gggaaagagcc at 82

40 <210> 89
<211> 74
<212> DNA
<213> Artificial Sequence

45 <220>
<223> Description of Artificial Sequence: primer nifD3

<400> 89
gatcagctgt tgaaagctat taaaccacaa aaaggattac tccggccctt atcacggta 60
cgacggattt gcta 74

50 <210> 90
<211> 74
<212> DNA
<213> Artificial Sequence

55 <220>
<223> Description of Artificial Sequence: primer nifD4

<400> 90
gatctagcaa atccgtcgta accgtgataa gggccggagt aatcctttt gtggtttaat 60
agctttcaac agct 74

5 <210> 91
 <211> 41
 <212> DNA
 <213> Artificial Sequence

10 <220>
 <223> Description of Artificial Sequence: primerSSV1-1

10 <400> 91
 ataagaatgc ggccgcccga tatgacgaaa gataagacgc g 41

15 <210> 92
 <211> 70
 <212> DNA
 <213> Artificial Sequence

20 <220>
 <223> Description of Artificial Sequence: primer SSV1-2

20 <400> 92
 ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgacgaaaag ataagacgcg 60
 ttataaaatac 70

25

30 <210> 93
 <211> 47
 <212> DNA
 <213> Artificial Sequence

35 <220>
 <223> Description of Artificial Sequence: primer SSV2

35 <400> 93
 tgtcccgggc tcgaaaccgg ggggatccgc ttgtagggga gtatccc 47

40 <210> 94
 <211> 47
 <212> DNA
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: primer SSV3

45 <400> 94
 gagcccgaaa caagcggaaag cggtggtgga aaagagggaa ctgaacg 47

50

55 <210> 95
 <211> 33
 <212> DNA
 <213> Artificial Sequence

55 <220>
 <223> Description of Artificial Sequence: primer SSV4

60 <400> 95
 atcgctcgag tcagaccct tttagccatt ccg 33

65 <210> 96
 <211> 40
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: primer SSV5
 5 <400> 96
 atcgttcgaa ggatccgcgt taacaccta gaaggcgaag 40

10 <210> 97
 <211> 38
 <212> DNA
 <213> Artificial Sequence

15 <220>
 <223> Description of Artificial Sequence: primer SSV6
 16 <400> 97
 atcgttcgaa ggatccgatt taatctctac atggagag 38

20 <210> 98
 <211> 64
 <212> DNA
 <213> Artificial Sequence

25 <220>
 <223> Description of Artificial Sequence: primer C31-2
 26 <400> 98
 ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgacacaag gggttgtgac 60
 30 cggg 64

35 <210> 99
 <211> 4831
 <212> DNA
 <213> Artificial Sequence

40 <220>
 <223> Description of Artificial Sequence: vector pRK41
 41 <400> 99
 ggccgcacca tgcccaagaa gaagaggaag gtgacacaag gggttgtgac cgggggtggac 60
 acgtacgcgg gtgcttacga ccgtcagtgc cgcgagcgcg agaattcgg 120
 45 ccagcgacac agcgtacgcg caacgaagac aaggcggccg accttcagcg cgaagtgcgag 180
 cgcgacgggg gcccgttcag gttcgtcggg catttcagcg aagcgcggg cacgtcggc 240
 ttcgggacgg cggagcgcgc ggagttcgg 1300 cgcacgcgcg 300
 ctcaacatga tcattgtcta tgacgtgcg cgcttctcgc gcctgaaggt catggacgcg 360
 attccgattt tctcggatt gtcgcctg ggcgtgacga ttgtttccac tcaggaaggc 420
 gtcttccggc agggaaacgt catggacctg attcacctga ttatgcggct cgacgcgtc 480
 50 cacaagaat ttgcgtgaa gtcggcgaag atttcgcaca cgaagaacct tcagcgcgaa 540
 tggggcggtt acgtcggcgg 1400 gaaggcgcct tacggcttcg agcttgcgtt 600
 gagatcagc gcaacggccg aatggtcaat gtcgtcatca acaagcttgc gcactcgacc 660
 actccccccta ccggaccctt cgagttcgg 1500 cccgacgtaa tccgggtgtg gtggcgtgag 720
 atcaagacgc acaaaccact tcccttcaag ccgggcagtc aagccgcccatt 1600 tcaccgggc 780
 55 agcatcacgg ggcttgcgtt ggcgttgcgtt gtcgtacgcgcg 1700 tgccgaccgc 840
 attggggaaa agacccgtt aagccgcctt 1800 aagccgccttgg gacccggcaaa ccgttgcgttgcg 900
 gacccgcgtt ttggggcgtt 1900 aatcccttcgg 1900
 60 ccgaccacga agattgaggg ttaccgcatt cagcgcgacc cgatcacgtt 2000 cccggccggc 1020
 gagcttgcgtt gccggaccgt 2100 catcgagccc gtcgtatgtt 2100 atgagcttca 2200
 gacggcaggg ggcgcggcaaa 2300 ggggcatttcc 2300 cgttgcgttgc 2400 cgccatggac 1140
 aagctgtact gcgagtgtgg 2500 cgcgcgttgc 2500 aatcccttcgg 2600
 gactcttacc gctgccgtcg 2700 cccgaagggt 2700 gtcgtacgc 2800 ccgcacctgg 2900
 ggcacgtgca acgtcagcat 2900 ggcggcactc 2900 gacaaggatcgc 3000 ttgcggaaacg 3100
 65 aagatcaggc acgccgaagg 3200 cgcgcactc 3200 gacaaggatcgc 3300 catcttcaac 3400
 cgcttcggca agctcactga 3500 ggcgccttgc 3500 aagagccggc 3600 aacggggcga 3700
 gagcgcgcgcg acgcctgaa 3800 cgcgccttgc 3800 gagctgtacg 3900 aagaccgcgc 4000
 tacgcacggac ccgttggcag 4100 gaaacacttc 4100 cggaaagcaac 4200 aggcagcgc 4300
 4500 gacgcacttc 4500 cggaaagcaac 4600 aggcagcgc 4700 gacgcgtccgg 4800

5 cagcaagggg cggaagagcg gcttgcgaa cttgaagccg ccgaagcccc gaagcttcc 1620
 cttgaccaat ggttcccccga agacccgcac gctgaccgcg ccggccctaa gtcgtgggtgg 1680
 gggcgccggt cagtagacga caagcgctg ttctgcggc tcttcgtaga caagatcggt 1740
 gtcacgaagt cgactacggg cagggggcag ggaacgccc tcgagaagcg cgcttcgatc 1800
 acgtggcga agccgcccac cgacgacgac gaagacgacg cccaggacgg cacggaagac 1860
 gtagccggt agccggccgtc cttagaactag tggatcccc gggctgcagg aattcgatat 1920
 caagcttac gataccgtcg acctcgaggg gggcccggt acccagctt tttccctt 1980
 atggagggtt aatttcgagc ttggcgtaat catggtcata gctgttctt gtgtgaaatt 2040
 10 gttatccgt cacaattcca cacaacatac gagccggaag cataaaagtgt aaagcctggg 2100
 gtgcctaattg agtgagctaa ctcacattaa ttgcgttgcg ctcaactccc gctttccagt 2160
 cgggaaaccc gtcgtccag ctgcattaaat gaatcgccca acgcgcgggg agaggccggt 2220
 tgcgttcttgc ggccttccgc gcttcctcgc tcactgactc gtcgcgtcg gtcgttccgc 2280
 tgcggcgaggc ggtatcagtc cactcaaagg cgtaatacgc gttatccaca gaatcagggg 2340
 ataacgcagg aaagaacatcg tgagaaaaaa gccgcaaaa ggccagaac cgtaaaaagg 2400
 15 ccgcgttgcg ggcgttttc cataggctcc gccccctga cgagcatcac aaaaatcgac 2460
 gtcagaatca gaggtggcga aacccgacag gactataaag ataccaggcg ttcccccctg 2520
 gaagctccct cgtgcgtct cctgttccga ccctgcgtt taccggatac ctgtccgcct 2580
 ttctcccttc gggaaacgtg ggcgttttc atagctcactc ctgttaggtat ctcagttcgg 2640
 20 tgttaggtcg tgcgttccaag ctgggctgtg tgacacgaa cccgcgttccgc cccgaccgt 2700
 ggccttatac cggtaactat cgttgcgtt ccaaccggc aagacacgac ttatcgccac 2760
 tggcagcaggc cactggtaac aggattagca gagcgaggta tgcgttccgc tgcgttccgc 2820
 tcttgaagtg gtggcctaac tacggctaca cttagaaggac agtatttggt atctgcgtc 2880
 tgcgttccgc agttacccgc ggaaaaaaggat ttgttagtcc tgcgttccgc aacaaacca 2940
 25 ccgcgttgcg cgggtgtttttt tttgttttca agcagcagat tgcgttccgc aacaaactcac 3060
 ctcaagaaga tcctttgatc ttttctacgg ggtctgcgc tgcgttccgc tgcgttccgc 3120
 gttaaaggat tttggtcatg agattatcaa aaaggatctt cacctagatc ctgtccgc 3180
 aaaaatgaag ttttaatca atctaaagta tatatgagta aacttggtct gacagttacc 3240
 aatgtttaat cagtggggca cctatctcgtc cgtatctgtct atttcggtca tccatagttg 3300
 30 cctgactccc cgtcggttag ataaactacga tacgggaggat cttaccatctt ggcggccagtg 3360
 ctgcaatgtat aacggcggagcc acacgctcac cgggtccaga tttatcagca ataaaccaggc 3420
 cagccggaaag ggccggagcgc agaagtggtc ctgcgttccgc atccgcgttca 3480
 ttaatttggt ccggggaaatc agagtaagta gttcgccagt tgcgttccgc tgcgttccgc 3540
 ttggcattgc tacaggcatc gtgggtcactc gtcgttccgc tgcgttccgc tgcgttccgc 3600
 35 ccgggttccca acgatcaagg cgagttacat gatccccat gttgtcaaa aaagcggtt 3660
 gtccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcgttccgc tgcgttccgc 3720
 ttatggcaggc actgcataat tctcttactg tcatgccatc gtcgttccgc tgcgttccgc 3780
 ctgggtgagta ctcaaccaag tcattctgag aatagtgtat aactttaaaa gtgtccatca 3840
 gccccggcgtc aatacgggat aataccggcgc cacatagcag accgcgttgc agatccagtt 3900
 40 ttggaaaacg ttcttcgggg cggaaatctt caaggatctt ttttacttcc accagcggtt 3960
 cgtatgttacc cactcggtca cccaaactgtat cttcagcatc gggaaataagg gcgcacacgg 4020
 ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaaa aagcatttat cagggttatt 4080
 aatgttgaat actcataactc ttcccttttc aatattatttgc gtcgttccgc tgcgttccgc 4140
 45 gtcgttccgc cggatatac ttgttaatgtat tttagaaaaaa aatattttgt taaaattcgc 4200
 gcaacatttc cggaaatgtt ccacctaaat tgcgttccgc gtcgttccgc tgcgttccgc 4260
 gtcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc 4320
 gtcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc 4380
 50 gtcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc tgcgttccgc 4440
 actaaatcg aaccctaaag ggagcccccc atttagagct tgacggggaa agccggcgaa 4500
 cgtggcgaga aaggaaggga agaaagcgaa aggagccggc gtcgttccgc tgcgttccgc 4560
 agccgtcactc ctgcgttccgc ccaccacacc cgcgcgtt aatgcgttccgc tacaggccgc 4620
 gtcgttccgc ccattcgatc tgcgttccgc ttggaaaggat gtcgttccgc tgcgttccgc 4680
 gtcgttccgc cagctggcga aagggggat gtcgttccgc tgcgttccgc tgcgttccgc 4740
 55 aggttttcc cagtcacgac ttgtaaaac gacggccagt gaattgtat acgactcact 4800
 ataggcgaa ttggagctcc accgcgggtgg c 4831

60 <210> 100
 <211> 20
 <212> DNA
 <213> Artificial Sequence

65 <220>
 <223> Description of Artificial Sequence: primer
 C31-screen 1

<400> 100

5 gcgtgagatc aagacgcaca

10 <210> 101
 <211> 21
 <212> DNA
 <213> Artificial Sequence

15 <220>
 <223> Description of Artificial Sequence: primer
 C31-screen 2

20 <400> 101
 gcagcggtaa gagtccttga t

25 <210> 102
 <211> 20
 <212> DNA
 <213> Artificial Sequence

30 <220>
 <223> Description of Artificial Sequence: primer
 beta-Gal 3

35 <400> 102
 atcctctgca tggtcaggtc

40 <210> 103
 <211> 18
 <212> DNA
 <213> Artificial Sequence

45 <220>
 <223> Description of Artificial Sequence: primer
 beta-Gal 4

50 <400> 103
 cgtggcctgca ttcattcc

55 <210> 104
 <211> 5878
 <212> DNA
 <213> Artificial Sequence

60 <220>
 <223> Description of Artificial Sequence: vector
 pCAGGS-Cre-pA

65 <400> 104
 cgccgcgtgc gccccgcgtc tggcggcgct gggggggcggt cgggggggtt 60
 tggcgctcc gctgtgcgc gggggggcg cggccgggggg cgggtccccgg cgggtgcgggg 120
 gggcgtgcgag gggaaacaaag gctgcgtgcg ggggtgtgc gtgggggggt gggcagggggg 180
 tggggcgccg ggggtcgccg tggtaacccccc ccctgcaccc ccctccccgg gttgctgagc 240
 acggcccgcc ttcgggtgcg gggctccgtg cggggcggtgg cggggggctc gccgtgcggg 300
 gcgggggggtg cggccagggtg ggggtgcggg cgggggggggg gccgcctcgg gccggggagg 360
 gtcggggggga gggggcgccg gggccggag cggccgggggg tggcgaggcg cggcgaggccg 420
 cagccattgc cttttatggt aatcggtgcga gggggcgccg ggactttctt tggccaaat 480
 ctggcgccgac cggaaatctgg gggccggccg cggcccccctt ctggcgccg cggggcgaaac 540
 ggtgccggccg cggcaggaaag gaaatggggcg gggggggctt tggcggtcg cggccggccc 600
 gtcccccttcc ccatctcccg cctcgccgtt cggccgggggg gacgggtgc ttcgggggggg 660
 acggggccagg ggggggttcg ctgtctggcg tggaccggc ggctcttagta agcgttgggg 720
 tggatactcc ctctaaaaag cgggcattgac ttctgcgtca agatgttcg tttccaaaaaa 780
 cgaggaggat ttgatattca cctggcccgcc ggtgtatgcctt tgggggtgg cccgcgtccat 840
 ctggtcagaa aagacaatct tttgttgtc aagcttgagg tggcgaggc ttgagatctg 900

5	gccatacact	tgagtgcacat	tgacatccac	tttgccttc	tctccacagg	tgtccactcc	960
	cagggcggcc	tgcaccatgc	ccaagaagaa	gaggaagggt	tccaatttc	tgaccgtaca	1020
	ccaaaatttgc	cctgcattac	cggtcgatgc	aaccgagtat	gagggtcgca	agaacctgtat	1080
	gacatgttc	agggatcgcc	aggcgtttc	tgagcatacc	tggaaaatgc	ttctgtccgt	1140
	ttgcccgtcg	tggccggcat	ggtgcaagtt	gaataaccgg	aatggtttcc	ccgcagaacc	1200
	tgaagatgtt	cgcattatc	ttctataatct	tcaggcgcgc	ggtctggca	taaaaactat	1260
	ccagcaacat	ttggccagc	taaacatgct	tcatcgtcg	tccgggctgc	cacgaccaag	1320
	tgacagcaat	gcttttcac	tggttatgct	gcccgtccga	aaagaaaaacg	ttgatgcagg	1380
10	tgaacgtgca	aaacaggctc	tagcgttgc	accgactgtat	ttcgaccagg	ttcgttca	1440
	catggaaaat	agcgatcgct	gccaggatat	acgtaatctg	gcatttctgg	ggattgctta	1500
	taacaccctg	ttacgtatag	cggaaattgc	caggatcagg	gttaaaagata	tctcacgtac	1560
	tgacgggtgg	agaatgtta	tccatattgg	cagaacgaaa	acgctggtt	gcaccgcagg	1620
	tgttagagaag	gcacttagcc	tgggggtaac	taaactggtc	gagcgtatgga	tttgcgtctc	1680
	ttgtgttagct	gatgatccga	ataactacct	gttttgcgg	gtcagaaaaaa	atggtgtgc	1740
	cgcgcacatct	gccaccaggc	agctatcac	tcgcgcctgt	gaagggattt	ttgaagcaac	1800
	tcatcgatgt	attiacggcg	ctaaggatga	ctctggtcag	agatacctgg	cctggctctg	1860
	acacagtgc	cgtgtcgagg	ccgcgcgaga	tatggcccg	gctggagttt	caataccgg	1920
	gatcatgca	gctgtggct	ggaccaatgt	aaatattgtc	atgaactata	tccgtaacct	1980
	gatagtagtga	acaggggcaa	ttgtgcgcct	gtcggagat	ggcgattagc	cattaacgcg	2040
20	taaatgatttgc	cagatccact	agttctaggg	ccgcgtcgc	ctcgagatcc	aggcgcggat	2100
	caataaaaaga	tcattat	caatagatct	gtgtgttgg	tttttgtgt	ccttggggga	2160
	gggggaggcc	agaatgaggc	ggggccaaagg	gggaggggg	ggccagaatg	acccctgggg	2220
	agggggaggc	cagaatgacc	ttgggggagg	gggagggccag	aatgaggcgc	gccccgggt	2280
	accgagctcg	aattcactgg	ccgtcg	acaacgtcg	gactggaaa	acccctggcg	2340
	tacccaactt	aatcgccctt	cagcacatcc	cccttgc	agctggcgta	atagcgaaga	2400
	gccccgcacc	gatgc	cccaacagtt	gogcagcctg	aatggcgaat	ggcgctgtat	2460
	gcggtatttt	ctc	atctgtcg	tat	tcacatgtt	gcactctcg	2520
	tacaatctgc	tctgatgccc	catagtaag	ccagcccc	cacccgcca	cacccgctg	2580
30	cgcgcctg	cgggcttgc	tgctccggc	atccgcttac	agacaagctg	tgaccgtctc	2640
	cgggagctgc	atgtgtcaga	gtt	gtc	aaacgcgcg	gacgaaaggg	2700
	cctcggtata	cgcc	ata	cat	aatgggtt	cttagacgtc	2760
	aggtggca	ttt	at	tg	ttt	tctaaataca	2820
	tccaaatatg	tatccgctca	tgagacaata	accctgataa	atgcttca	aatattggaa	2880
	aggaagagat	atgatgtt	aa	at	ttt	ttg	2940
35	ttgccttct	gttttgc	acc	ca	gtt	ctg	3000
	gttgggtgca	cgagtgggtt	cc	at	gg	tc	3060
	tttgc	gaagaacgtt	at	ca	gg	tt	3120
	gttattatcc	cgat	cc	gg	gg	gg	3180
	gaatgacttgc	gtt	acc	gg	gg	gg	3240
40	aagagaatttgc	tgc	cc	at	gg	gg	3300
	gacaacgatc	cg	ata	gt	gg	gg	3360
	aactcgcc	gg	acc	gt	gg	gg	3420
	caccacgat	gg	gg	gt	gg	gg	3480
	tactctagct	cc	ttt	gg	gg	gg	3540
45	acttctgc	cc	ttt	gg	gg	gg	3600
	gcgtgggt	cc	ttt	gg	gg	gg	3660
	agttatctac	cc	ttt	gg	gg	gg	3720
	acgacgggg	cc	ttt	gg	gg	gg	3780
	gataggtg	cc	ttt	gg	gg	gg	3840
50	ttagattgtat	tt	ttt	gg	gg	gg	3900
	taatctcat	tt	ttt	gg	gg	gg	3960
	aaaaaaagatc	tt	ttt	gg	gg	gg	4020
	acaaaaaaa	tt	ttt	gg	gg	gg	4080
	ccaccgctac	tt	ttt	gg	gg	gg	4140
55	tttccgaag	tt	ttt	gg	gg	gg	4200
	gtactgttgc	tt	ttt	gg	gg	gg	4260
	ggcgtagtt	tt	ttt	gg	gg	gg	4320
	atccgttgc	tt	ttt	gg	gg	gg	4380
	aagacgtat	tt	ttt	gg	gg	gg	4440
60	ggccagcttgc	tt	ttt	gg	gg	gg	4500
	gaggcaacga	tt	ttt	gg	gg	gg	4560
	aagcgcac	tt	ttt	gg	gg	gg	4620
	acaggagag	tt	ttt	gg	gg	gg	4680
	cgcacgagg	tt	ttt	gg	gg	gg	4740
	cggttgc	tt	ttt	gg	gg	gg	4800
	cac	tt	ttt	gg	gg	gg	4860
65	cctatggaaa	tt	ttt	gg	gg	gg	4920
	tgctcacat	tt	ttt	gg	gg	gg	4980
	tgatgttagct	tt	ttt	gg	gg	gg	5040
	ggaagcg	tt	ttt	gg	gg	gg	5100
	atgcagctgg	tt	ttt	gg	gg	gg	5160
	cac	tt	ttt	gg	gg	gg	5220
	tgttagtt	tt	ttt	gg	gg	gg	5280
	ctca	tt	ttt	gg	gg	gg	5340
	gtt	tt	ttt	gg	gg	gg	5400

cgccaagcta	gccccggcta	gcttgcata	ctgcagggtt	tcgacattga	ttattgacta	4980	
gttattaata	gtaatcaatt	acggggtcat	tagttcatag	cccatatatg	gagttcccg	5040	
ttacataact	tacggtaat	ggcccgctg	gctgaccgc	caacgacccc	cgcccattga	5100	
5	cgtcaataat	gacgtatgtt	cccatagtaa	cgccaatagg	gactttccat	tgacgtcaat	5160
gggtggacta	tttacggtaa	actgcccact	tggcaigtaca	tcaagtgtat	catacgccaa	5220	
gtacgcccc	tattgacgtc	aatgacggta	aatggcccgc	ctggcattat	gcccagtaca	5280	
tgacctttag	ggactttct	actggcagt	acatctacgt	attagtcata	gctattacca	5340	
tggtcgagg	tgagccccac	gttctgttc	actctcccc	tctccccccc	ctccccaccc	5400	
10	ccaattttgt	atttattttat	tttttaatta	tttgtgcag	cgatggggc	gggggggggg	5460
ggggcgccgc	ccaggcgffff	cggggcgggg	cgagggggcgg	ggcggggcga	ggcggagagg	5520	
tgccgcggca	gccaatcaga	gccccgcgcgt	ccggaaagttt	ccctttatgg	cgaggcgccg	5580	
gcggcgccgg	ccctataaaaa	agcgaagcgc	gccccggggcgc	ggagtcgtc	cggtgccttc	5640	
gccccgtgcc	ccgcctccgcg	ccgcctcgcg	ccgccccccc	ccgctctgac	tgaccgcgtt	5700	
15	actccccacag	gtgagccggc	gggacggccc	tttctctccg	ggctgttaatt	agcgtttgtt	5760
ttaatgacgg	ctcggttctt	ttctgtggct	gcgtgaaaggc	cttaaagggc	tccgggaggg	5820	
cccttgc	ggggggggagc	ggctcgffff	gtgcgtgcgt	gtgtgtgtc	gtggggag	5878	

20 <210> 105

20 <211> 6641

20 <212> DNA

20 <213> Artificial Sequence

25 <220>

25 <223> Description of Artificial Sequence: vector
pCAGGSC31CNLS-pA

30 <400> 105

attgatttatt	gactagttat	taatagtaat	caattacggg	gtcatttagtt	catagcccat	60	
atatggggat	ccgcgttaca	taacttacgg	taaatggccc	gcctggctga	ccgccccaaacg	120	
accccccggcc	attgacgtca	ataatgaacgt	atgttcccat	agtaacgcga	atagggactt	180	
tccattgacg	tcaatgggt	gactattac	ggtaaaactgc	ccacttggca	gtacatcaag	240	
tgtatcatat	gccaagtacg	ccccctattt	acgtcaatga	cgtaaaatgg	ccgcctggc	300	
35	attatgcccc	gtacatgacc	ttatggact	ttctctacttgc	gcagttacatc	tacgtattag	360
tcatcgctat	taccatgggt	cgaggtgagc	cccacgttct	gcttcaactt	ccccatctcc	420	
ccccctccccc	caccccaat	tttgcattttt	tttgcattttt	aattattttt	tgcaagcgatg	480	
ggggcgggggg	ggggggggggc	ggggcgccagg	cgggggcgffff	cgggggcgagg	ggcgccccgg	540	
ggcgaggcg	agaggtgcgg	ccgcagccaa	tcaagacggc	cgctccgaa	attttcctt	600	
40	tatggcgagg	oggccggggc	ggggcccta	taaaaagcga	agcgcgcggc	gggcgggagtt	660
cgctcggttgc	ctttcgcccc	gtgccccgt	ccgcgcggcc	tcgcgcggcc	cgcccccggt	720	
ctgactgacc	gcgttactcc	cacagggtgag	cgggcgggac	ggcccttctc	ctccgggctg	780	
taatttagcgc	ttgggtttaat	gacggctctgt	ttttttctgt	tggctgcgt	aaagccttaa	840	
45	agggtccccc	gaggggccctt	tgtgcgggg	ggagcggcgc	gggggggtgc	tgcgtgtgt	900
tgtgcgtgg	gagcgcgcgc	tgcggccccc	gctgtgtgc	gctgtgtgc	gtgcggggcgc	960	
ggcgccgggg	ctttgtgcgc	tccgcgtgt	cgcgaggggg	gcccggccgg	ggcgggtggcc	1020	
ccgcgggtgc	ggggggctgc	gaggggaaaca	aaggtgcgt	gggggggtgc	tgcgtggggg	1080	
50	ggtagcagg	gggtgtgggc	ggggcggtcg	gggtgtaa	ccccccatccc	1140	
cgagttgcgt	agcacggccc	ggcttcgggt	cgccccggcc	gtgcggggcc	tgcgcgggg	1200	
ctgcgcgtgc	cgggcggggg	gtggcgccag	gtgggggtgc	cgggcgggcc	ggggccgcgt	1260	
cgggccgggg	agggtctcggg	ggagggggcgc	ggggggcccg	gagcgcggc	ggctgtcgag	1320	
55	gcccggcgag	ccgcagccat	tgcctttat	gttaatcgtg	cgagaggggc	cagggacttc	1380
ctttgtccca	aatctggcg	agccgaatc	tggaggcgc	ccgcgcaccc	cctctagcgg	1440	
gcccggcg	agcgggtgcgg	cccccggcagg	aaggaaatgg	gccccggagg	ccttcgtcgc	1500	
tcgcgcgc	gccgtcccc	tctccatctc	cagcctcggt	gtgcgcgcag	ggggacgggt	1560	
60	gccttcgggg	gggggggtt	gggggggtt	gggtgtgacc	ggcggtctct	1620	
gtaagcggtt	gggtgagttac	tccctctcaa	aaggggcat	gacttctgcg	ctaagattgt	1680	
cagttccaa	aaacgaggag	gatttgat	tcacctggcc	cgcggtgtat	cctttgaggg	1740	
tggccgcgtc	catctggtca	aaaaagacaa	tctttttgtt	gtcaagcttgc	agggtgtggca	1800	
65	ggcttggat	ctggccatac	acttggatgt	cattgacato	cactttgcct	ttctctccac	1860
agggttccac	tcccaggggcg	ggcccccgt	atgacacaag	gggttgcac	cggggtggac	1920	
acgtacgcgg	gtgcgttacga	ccgtcagtgc	cgcgagcgcg	agaattcgc	cgcaagcaagc	1980	
ccagcgacac	agcgtagcgc	caacgaagac	aaggcgccgc	accttcagcg	cgaagtgcag	2040	
cgcgacgggg	ggcggttcag	gttcgtcggg	catttcagcg	aaggcgccgg	cacgtcgggc	2100	
ttcgggacgg	cgagacggccc	ggagttcga	cgcatctgc	acgaatcgc	cgccggggcgg	2160	
ctcaacatga	tcatgtctt	tgacgtgtcg	cgcttcgc	gcctgaaggt	catggacgcg	2220	
attccgattt	tctcggaaat	gtcgcctctg	ggcgtgacga	ttgtttccac	tcaggaaggc	2280	
gtcttcggc	agggaaacgt	catggacctg	attcacatga	ttatgcggct	cgacgcgtcg	2340	

	cacaaaagaat	cttcgctgaa	gtcggcgaag	attctcgaca	cgaagaacct	tcagcgcgaa	2400
	ttgggcgggt	acgtccggcg	gaaggcgctt	tacggcttcg	agcttggttc	ggagacgaaag	2460
5	gagatcacgc	gcaacggccg	aatggtcaat	gtcgctatca	acaagcttgc	gcactcgacc	2520
	actccccc	ccggaccctt	cgagttcgag	cccgacgtaa	tccgggtgt	gtggcgtgag	2580
	atcaagacgc	acaaacacct	tcccttcaag	ccgggcagtc	aagccgccc	tcacccgggc	2640
	agcatcacgg	ggcttgtaa	gcccgcgtt	gctgacgccc	tgccgacc	ggcgagacg	2700
	attgggaaga	agaccgctt	aaggccttgg	gacccggcaa	ccgttatgcg	aatccttcgg	2760
10	gaccgcgtt	ttgcgggtt	cgccgcgtt	gtgatctaca	agaagaagcc	ggacggcacg	2820
	ccgaccacga	agattgaggg	ttaccgcatt	cagcgcgacc	cgatcacgt	ccggccgggtc	2880
	gagttgatt	gcggaccgt	catcgagccc	gctgagtgg	atgagctca	ggcgtgggt	2940
	gacggcagg	ggcgcggcaa	ggggcttcc	cgggggcaag	ccattctgtc	cgccatggac	3000
	aagctgtact	gcgagtgtgg	cgccgtcat	acttcgaagc	gccccggaa	atcgatcaag	3060
15	gactcttacc	gtgcggctcg	ccggaaaggt	gtcgaccctgt	ccgcaccc	gcagcacgaa	3120
	gcaacgtgca	acgtcagcat	ggcggcaact	gacaaggatcg	ttgcggaa	catcttcaac	3180
	agatcaggc	acgcggaaagg	cgacggaaag	acgttggcgc	ttctgtggga	agccggcccg	3240
	cgcttcggca	agetcactgt	gggcgcgtt	aagagcggcg	aacggggcga	ccttggcgcg	3300
	gagcgcgcg	acgccttgc	cgcccttgc	gagctgtacg	aagaccgcgc	ggcaggcgcg	3360
20	tacgacggac	ccgttggcag	gaagcacttc	cggaagcaac	aggcagcgt	gacgctccgg	3420
	cagcaagggg	cggaagagcg	gcttgcggaa	cttgaagccg	ccgaagcccc	gaagcttccc	3480
	cttgaccat	ggttccccga	agacgcgcac	gctgaccga	ccggccctaa	gtcgtgggt	3540
	gggcgcgcgt	cagtagacga	caagcgcgt	ttctgtggc	tcttcgtaga	caagatcg	3600
	gtcacgaat	cgactacggg	cagggggcag	ggaacgccc	tcgagaagcg	cgcttcgatc	3660
25	acgtggcga	agccggcgcac	cgacgacgac	gaagacgacg	cccaggacgg	cacggaagac	3720
	gtagcggcgc	ctaagaagaa	gaggaaggtt	tagactctcg	agatccaggc	gcggatcaat	3780
	aaaagatcat	tattttcaat	agatctgtgt	tttggttttt	tgtgtgcctt	gggggagggg	3840
	gaggccagaa	tgaggcgcgg	ccaaggggga	gggggaggcc	agaatgacct	tgggggagggg	3900
	ggaggccaga	atgacottgg	gggaggggg	ggccagaatg	aggcgcgc	ccgggtaccc	3960
30	agctcgaatt	cactggccgt	cgitttacaa	cgtcgtact	gggaaaaccc	tggcgttacc	4020
	caacttaatc	gccttgcagc	acatccccct	ttcgcctagct	ggcgttaatag	cgaagaggcc	4080
	cgcacccgatc	gccccttccca	acatggcgc	agctgaatg	gcaatggcg	cctgatgcgg	4140
	tattttctcc	ttacgcatt	gtgcgtt	tcacccgc	tatggtgcac	tctcgtatca	4200
	atctgtctg	atgcgcata	gttaaggccag	ccccgacac	cgccaacacc	cgctgacgcg	4260
	ccctgacggg	cttgcgtct	ccggcatacc	gcttacagac	aagctgtac	cgctccggg	4320
35	agctgcatgt	gtcagagggt	tttcaccgtca	tcacccgaaac	gcmcagacg	aaaggccctc	4380
	gtgatacggc	tattttata	gtttaatgtc	atgataataa	tggtttctta	gacgtcagg	4440
	gcaacttttc	ggggaaatgt	gcmcggaaacc	cctattttgtt	tattttcta	aatacattca	4500
	aatatgtatc	cgctcatgag	acaataaccc	tgataaaatgc	ttcaataata	ttgaaaaagg	4560
	aagagtatga	gtattcaaca	tttccgtgtc	gccttatttgc	ccttttttgc	ggcattttgc	4620
40	tttcctgttt	ttgctcaccc	agaaacgctg	gtgaaagtaa	aagatgtga	agatcagtgt	4680
	gtgtcacgag	ttgggttacat	cgaaactggat	ctcaacagcg	gtaagatcct	tgagagttt	4740
	cgcggccaaag	aacgttttcc	aatatgtatc	acttttaaag	ttctgtatg	tggcgcggta	4800
	ttatcccgt	ttgacgcgg	gcaagagca	ctcggtcgcc	gcatacacta	ttctcagaat	4860
45	gacttggtt	agtaactcacc	agtcacagaa	aagcatctta	cggtatggcat	gacagtaaga	4920
	gaatttatgca	gtgcgtccat	aaccatgagt	gataacactg	cgggcaactt	acttctgaca	4980
	acgtatcgag	gaccgaagga	gctaaccgct	tttttgcaca	acatggggga	tcatgtact	5040
	cgccttgcata	gttggaaacc	ggagctgaat	gaaccatata	caaacgacg	gcgtgacacc	5100
	acgtatcgct	tagcatatgc	aacaacgtt	cgcacaaat	taacttgcg	actacttaat	5160
50	ctagcttccc	ggcaacataa	aatagactgg	atggaggccg	ataaagtgc	aggaccactt	5220
	ctgcgtctgg	cccttccggc	ttggctgttt	atttgcata	aatotggagc	cggtgagcgt	5280
	gggtctcgcg	gtatcattgc	agcaacttgg	ccagatggta	agccctcccg	tatctgtattt	5340
	atctacacga	cggggagtc	ggcaactat	gatgaacgaa	atagacagat	cgctgagata	5400
	ggtgcctcac	tgattaagca	ttggtaact	tcagaccaag	tttactcata	tataactttag	5460
55	attgatttaa	aacttcattt	ttaattttaa	aggatctagg	tgaagatcct	ttttataat	5520
	ctcatgacca	aaatccctta	acgtgagttt	tgcgttccact	gagcgtcaga	ccccgtagaa	5580
	aagatcaaag	gatcttcatttgc	agatcctttt	tttctgcgcg	taatctgtc	tttgcacaaaca	5640
	aaaaaaccac	cgctaccacgc	ggtgggtttt	ttgcggatc	aagagctacc	aacttctttt	5700
	ccgaaggtaa	ctggcattcag	cagagcgcag	ataccaaata	ctgtccctct	agtgtagccg	5760
	tagttagggc	accacttcaa	gaactctgt	gcacccgcata	cataccctcg	tctgctaattc	5820
60	ctgttaccag	tggctgtctgc	cagtggcgat	aagtcgtgtc	ttaccgggtt	ggactcaaga	5880
	cgatagttac	cggataaggc	gcagcggctg	ggctgaacgg	gggttctgt	cacacagccc	5940
	agcttggagc	gaacgaccta	caccgaact	agatacctac	agcgtgagct	atgagaaagc	6000
	gccacgcgtc	ccgaaaggggag	aaaggcggac	aggatccgg	taagcggcag	gtcggaaaca	6060
	ggagagcgc	cgaggggagct	tccaggggga	aacgccttgt	atctttatag	tccgtcg	6120
	tttccgcacc	tctgacttgc	gctgtcgat	ttgtatgtc	cgtcagggggg	gcccggccct	6180
65	tggaaaaaac	ccagcaacgc	ggccttttta	cggttccgtt	ccttttgcgt	gccttttgcgt	6240
	cacatgttct	ttccgtcgat	atccccctgt	tctgtggata	accgtattac	cgccctttgag	6300
	ttagctgata	ccgctcgccg	cagccgaacg	accgagcgc	cgagtcagt	gagcgaggaa	6360

```

gccaatatacg caaacccgcct ctcggcgat tcattatgc 6420
acagggttcc cgactggaaa gccccggatg agcgcaacgc aattatgtg 6480
atcttgcgtt ctcattaggc acggggggct ttacacttta tgcttcggc tcgtatgtt 6540
tgtggattt tggcgatata acaatttccac acaggaaaca gctatgacca tgattacgca 6600
aaggctggcc gggcttagtt gcatgcgtgc aggttttgc c 6641

```

10 <210> 106
<211> 11784
<212> DNA
<213> Artificial Sequence

15 <220>
<223> Description of Artificial Sequence: modified
ROSA26 locus

<400> 106													
20	ggcaggccct	ccgagcgtgg	tggagccgtt	ctgtgagaca	gccgggtacg	agtcgtgacg	60	cttggaaagggg	caagcgggtg	gtgggcagga	atgcggtccg	ccctgcagca	accggagggg
	gagggagaag	ggagcggaaa	agtctccacc	ggacgccc	atggctcggg	gggggggggg	180	cagcggagga	gcgctccgg	ccgacgttc	gtcgctgatt	ggcttctttt	cctcccgccc
	tgtgtgaaaa	cacaaatggc	gtgttttgg	tggcgtaagg	cgccctgtcg	ttaacggcag	300	ccggagtgcg	cagccgtccgg	cagcctcgct	ctgcccactg	gttggggcgg	gaggtaggtg
	gggtgaggcg	agctggacgt	gcggggcggg	tggccctctg	gcggggcggg	ggaggggagg	420	gagggtcago	gaaagttagct	cgcgcgogag	cgcccgcaca	ccctccccc	cctctggggg
25	agtcgtttt	cccgccgccc	gcccggcctc	gtcgctgtat	tggctctcg	ggcccagaaa	480	actggccctt	gccattggct	cgtgttctgt	caagttaggt	ccatccgccc	gcaagcgggg
	ggggcgagga	ggcgctccca	ggttccggcc	ctccccctcg	ccccggcg	cagagtctgg	660	cccgccgccc	ctgcccacgg	gttccggatc	ggcgctgg	ggggggggac	gggcgtatgg
30	cccgccgccc	ctgcccacgg	ttggcggag	cgcgcgctgg	ggggggggac	gggcgtatgg	720	gctgagcggc	tgcggggcg	gtgcaagcac	gttccgact	tgagttgcct	caagaggggc
	gtgctgagcc	agacccat	ttggggcttt	ttggaggcag	gaagcacttg	ggggagggc	780	ttggggcttt	ttggaggcag	cgcgcaactc	ggggactcg	gggggctcag	gagggctcag
	gtaaagggagc	tgcagtggag	taggcgggga	gaaggccgca	cccttcctcg	gagggggggag	960	gggggtgttg	caataccctt	tcccccctt	ccctcgtat	ctgcaactcc	agtccttgc
35	ctggggctgg	gagaatccct	cagggtgtac	ctgcacgtct	tatcctgtcc	cacagctcgc	1080	cttaggttacc	gatatccctg	ctggggatcc	tctgtgcct	cggccgttct	gaggaccgccc
	ctaggttaacc	gatatccctg	ttccttgcgt	atgtcatact	aaactcttcg	ctggggatcc	1140	tttggggctgg	tttggggctgg	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	aaactcttcg	cggtctttcc	gagaccaact	tcactaacaa	cgcccgggga	cccccggcc	1200	gagaccaact	tcactaacaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
40	cgcccgggga	gccccagggc	ataacttcgt	atagcataca	cgcccgggga	cccccggcc	1260	cgcccgggga	gatgggttag	ccggcttcgg	accgtcacga	gtggagtaacg	1320
	gagggcgctt	ttcccaaggc	cgctacacaa	gtggccctcg	tttggggctgg	cccccggcc	1380	tttggggctgg	tttggggctgg	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	tttggggctgg	tttggggctgg	tttggggctgg	tttggggctgg	tttggggctgg	tttggggctgg	1440	tttggggctgg	tttggggctgg	tttggggctgg	tttggggctgg	tttggggctgg	tttggggctgg
45	ccccccccc	gcagctcgcg	ctccgttctt	tttggggctgg	tttggggctgg	tttggggctgg	1500	gtctcgatca	gatggacacgc	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	gtctcgatca	gatggacacgc	gtttgtctcc	gccaatagca	ggggcgggct	ggggcggggg	1560	gtttgtctcc	gccaatagca	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	gccaatagca	gtttgtctcc	ggggcgggct	caggggggg	gtttgtctcc	ggggcggggg	1620	gtttgtctcc	gatggacacgc	cccccggcc	cccccggcc	cccccggcc	cccccggcc
50	gtttgtctcc	acgcttcaaa	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	1680	gtttgtctcc	gatggacacgc	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	gagaagtttc	tgatcgaaaa	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	1740	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	gagaagtttc	tgatcgaaaa	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	1800	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
55	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	1860	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	1920	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	1980	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
60	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2040	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2100	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2160	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
65	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2220	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2280	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2340	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2400	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2460	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2520	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2580	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2640	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2700	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2760	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2820	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2880	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	2940	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc
	ggccctttcg	acgatccagc	ggccctttcg	acgatccagc	gagaagtttc	tgatcgaaaa	3000	ggccctttcg	tgatcgaaaa	cccccggcc	cccccggcc	cccccggcc	cccccggcc

acaataaaaga tgtccactaa aatggaaagtt ttccctgtca tactttgtta agaagggtga 3060
 gaacagagta cctacatttt gaatggaaagg atggagcta cgggggtggg ggtgggggtgg 3120
 gattagataa atgcctgctc ttactgaag gctctttact attgctttat gataatgtt 3180
 5 catagttgga tatacataatt taaacaagca aaaccaaatt aagggccagc tcattcctcc 3240
 cactcatgtat ctatacatct atagatctc cgtggatca ttgttttct cttgattccc 3300
 actttgtggt tctaagtact gtggtttcca aatgtgtcag ttcataagcc tgaagaacga 3360
 gatcagcagc ctctgttcca catacactc atttcagta ttgtttgcc aagttctaat 3420
 tccatcagaa gcttcagctg ctcgactaga ggatcataat cagccatacc acattttgtag 3480
 10 aggtttact tgctttaaaa aacctccac acctccccc acacccatggaa cataaaatga 3540
 atgcaattgt tggtttaac ttgttattt cagttataa tggttacaat taaagcaata 3600
 gcatcacaat tttcacaat aaagcattt ttcaactgca ttcttagttgt ggttigtcca 3660
 aactcatcaa tggatcttat catgtctgga tccgtgtcat gtcggcggacc ctacgcccc 3720
 aactgagaga actcaaaggt taccggactt gggcaactac tccgaaacac cgcttctgga 3780
 tccataactt cgtatagcat acattatacg aagttatacc gggccacat ggtcgcgagt 3840
 15 agcttggcac tggccgtcgt ttacaacgt cgtgactggg aaaacccctgg cgttacccaa 3900
 cttaatcgcc ttgcagcaca tcccccttc gcaagctggc gtaatagcga agaggccgc 3960
 accgatcgcc ctccccaca gttgcgcagc ctgaatggcg aatggcgtt tgcctggtt 4020
 cccgcaccag aagcggtgcc gggaaagctgg ctggagtgcg atcttcctga ggccgatact 4080
 20 gtcgtcgtcc cctcaaactg gcagatgcac ggttacatgcg cgtttatctca caccacgt 4140
 acctatccca ttacggtaa tcogccgtt gttcccacgg agaatccgac ggggtttag 4200
 tcgctcacat ttaatgttga taaaagctgg ctacagggaa gccagacgac aattatttt 4260
 gatggcgta actcgccgtt tcatctgtgg tgcacacggc gctgggtcgg ttacggccag 4320
 gacagtgtt tgccgtctga atttgcactt agcgcatttt tacgcgcgg agaaaaccgc 4380
 25 ctcgcggta tgggtctgcg ttggagtgcg ggcagttatc tggaaagatca ggatatgtgg 4440
 cggatgagcg gcatttccg tgacgtctcg ttgtgcata aaccgactac acaaattcagc 4500
 gatitccatg ttgcactcg cttaatgtat gatttcagcc gcgctgtact ggaggctgaa 4560
 gttcagatgt gcggcgagtt gctgtactac ctacgggtaa cagtttctt atggcagggt 4620
 gaaacgcagg tcgcaggcgg gttatgtccg caccgcgcct ttccggcgtt aaattatcga ttagcgtgg 4680
 30 gaaatccca atctctatcg tgcgggtgtt gaacttcacca cccgcaactac aaccgcgt 4740
 gaagcagaag ctgcgtatgt ctgcacccggc gatggcgaccc cccgcacgg cacgctgatt 4800
 ctgcacccgca agccgtgtt gatggcgaccc gatggatgtt ttgaaaatgg tctgtctgt 4860
 ggtcaggta tggatgagca gacgtatgtt cggatccgc acgagoatca tccctctgc 4920
 35 tttaaacccg tgcgtcttc gcatattatccg aaccatccgc tgggttacac gctgtgcac 4980
 cgcgtacggcc tggatgtgtt ggtgaagcc aatattgaaa cccacggcat ggtccaatg 5040
 aatcgtctga ccgatgtatcc ggcgtggcta ccggcgatga gcaacccgt aacgcgaatg 5100
 gtgcagcgcg atcgtatca cccgaggtgtt atcatctgtt cgctggggaa tgaatcaggc 5160
 40 cacggcgcta atcacgcgc gctgtatcg tggatcaat ctcgtatcc ttcccgcccc 5220
 gtgcagtatg aaggccggg ggcgtcgatcc gatggatgtt tggatcaat ctcgtatcc 5280
 ggcgcgtgg atgaagacca gcccggatcc gtcgtatcc atattatttgc 5340
 cttcgcgtac ctggagagac gcccggatcc gtcgtatcc aatggtccat caaaaaatgg 5400
 aacagtcttgcg ggggttccg taaatactgg caggcgatcc gtcgtatcc cctgtttacag 5460
 45 gggcgcttcg tctggactg ggtggatcg tcgtgttata aatatgtatgc aacacggcaac 5520
 ccgtggtcgg cttacggccgg tgattttggc gatacggccg acgatcgcca gttctgtatg 5580
 aacggcttcg tctttccg cccgacggcc catccagcgc tgacggaaacccaaacccag 5640
 cagcgtttt tccagttccg ttatccggg caaaccatcg aagtggccac cgaataccctg 5700
 50 ttcgcgtata gcgataacga gctctgcac tggatgtgg gctggatcg taagccgttg 5760
 gcaaggcggtt aagtgcctct ggtgtcgatcc cacaaggta aacagtgtatgc tggatgtt 5820
 gaactacccg acggcgagag cccggggaa ctctggctca cgtacgcgt agtcaacccg 5880
 aacgcgaccc catggtcaga cccggccgc tcccacgcca tcccgcacatc gaccaccagc 5940
 55 gaaaacccca gtgtgacgct cccggccgc tcccacgcca aattnaaccg ccagtcaggc 6000
 gaaatggatt tttgcatacg gtcggatataa aaccaactgc tgacccgcgt ggcgtatcg 6060
 tttcttcac agatgtggat tggcgataaa taacgacatt ggcgttaagt aagcgaccccg cattgaccct 6120
 ttcacccgtg caccgctggc taacgacatt ggcgttaagt aagcgaccccg cattgaccct 6180
 60 aacgcctggg tcgaacgcgtc gaaaggccgcg ggccattacc agggccaaacccggccatggc 6240
 cagtgcaccc cagatacact tgctgtatcg gtgtgttataa aattnaaccg ccagtcaggc 6300
 catcaggggaa aaacccatcg tatcagccgg aaaacccatcg ggtgtatcg tagtggtcaa 6360
 atggcgatata cccgtatgt tgaagtggcg agcgatcac cgcacccgc gcggattggc 6420
 ctgaactgcc agctggcgca ggtgcacccatcg ggggtaaact ggctggatcg agggccgcaa 6480
 gaaaactatc cccgaccgcct tactggcc tggatgttgc gctggatct gccattgtca 6540
 gacatgtata ccccgatcg ttcacccgtg gaaaacccgc tgcgtatcc gacgcgcgaa 6600
 ttgaattatg gcccacacca cggccatccg gacttccagt tcaacatcg cgcgtatcc 6660
 caacagcaac tgatggaaac cggccatccg catctgtgc acggaaacccggccatggc 6720
 ctgaatatacg acgggttccg tatggggatt ggtggcgaccc tccgtatcc cccgtatcg 6780
 tcggcgaaat tccagctgatcc taccattacc agtggatcg tggatgttgc 6840
 taataataac cggcgagggg ggtatccatcg gatgttgc tggatgttgc tggatgttgc 6900
 attggacaaa ctacccatcg agattnaaccg tccatccatcg aattnaaccg ttttaatgtt 6960
 attggacaaa ctacccatcg agattnaaccg tccatccatcg aattnaaccg ttttaatgtt 7020

ataatgtgtt aaactactga ttctaatgtt ttgtgtattt tagattccaa cctatggAAC 7080
 tggatggc gggcgtt gggatgcac atccagacat gataagatac attgtatgt 7140
 ttggacaaac cacaactaga atgcagtgtt aaaaatgtttt tattttgttattt attgtatgt 7200
 5 ctattgtttt atttgcattt gtttcagggtt cagggggagg tggggggagg tttttaaagc aagtaaaacc 7260
 ttcattttt gtttcagggtt gtttcagggtt cagggggagg tggggggagg tttttaaagc aagtaaaacc 7320
 tctacaaatg tggatgtt gtttcagggtt cagggggagg tggggggagg tttttaaagc aagtaaaacc 7380
 ggttaaccgaa gttccatatac tttcttagaga ataggaactt cggaaataggaa acttcaagct 7440
 taagcgctag cctagaagat gggcgggggtt cttctggca ggttaaccgaa aacaatttgc 7500
 10 gtgtgggggtt tggctgtcag gggatgtt cttctggca ggttaaccgaa aacaatttgc 7560
 cacagatttt cgggtttgtc gggatgtt ttaataggggg caaataagga aaatggggagg 7620
 ataggtagtc atctgggtt ttatgcagca aaactacagg ttattttgtc ttgtgtatccg 7680
 cctcgaggta ttttccatcg aggttagatta aagacatgtt caccgggtt ttttccatcg 7740
 ctgcttgaga tccttactac agttagaaat tacagtgtc cttctggca ggttaaccgaa aacaatttgc 7800
 gaatttttaat catttttaaa gagcccgta ctccatatccg 7860
 15 agccttatac aaaggatttt tagaacaactt atttgcattt ctttttgcatttattttactt 7920
 gtttatccaa ccccttagaca gggatgtt ctttttgcatttattttactt 7980
 tgactcatga aaccagacag attgttaca tacaccacaa atcgaggctg tagctggggc 8040
 ctcaacactg cagtttttataactt ctttttgcatttattttactt 8100
 20 tccttaattt tcagttgtcgtt tcacctctcc ctttttgcatttattttactt 8160
 ctcagtccag ggagttttac aacaatagat gtatttagaa tccaaacctt agcttaactt 8220
 tccactccca tgaatgcctc tttttttttt ctttttgcatttattttactt 8280
 aatgggttcca ggtggatgtc tttttttttt ctttttgcatttattttactt 8340
 ctgatattttt aagacattaa aaggtagtattt tcatttttgcatttattttactt 8400
 25 gtttactaaa attttgcatttattttactt ttttttgcatttattttactt 8460
 gttcagggtgtt ttgttgcatttattttactt ttttttgcatttattttactt 8520
 aaggcgtgtt ttgttgcatttattttactt ttttttgcatttattttactt 8580
 aaatgtgatt ttgttgcatttattttactt ttttttgcatttattttactt 8640
 caagtaaaat gattaagcaaa ctttttgcatttattttactt 8700
 30 gtgtgtgtttt ttgttgcatttattttactt ttttttgcatttattttactt 8760
 agagcacaga ctgttcttcc ttttttgcatttattttactt 8820
 cacaaccatc ttttttgcatttattttactt 8880
 attcacattt aataaataaaa ttttttgcatttattttactt 8940
 ttttttgcatttattttactt 8940
 35 ttttttgcatttattttactt 8940
 caataatcaa attactctt ttttttgcatttattttactt 9000
 agttaactg ttttttgcatttattttactt 9060
 agactttat ttttttgcatttattttactt 9120
 40 ttttttgcatttattttactt 9180
 atcctattttt ttttttgcatttattttactt 9240
 aaagctaagt ctgttgcatttattttactt 9300
 aaaatgttagg gtttttgcatttattttactt 9360
 ctttttgcatttattttactt 9360
 ctttttgcatttattttactt 9420
 45 ttttttgcatttattttactt 9420
 ttttttgcatttattttactt 9480
 ttttttgcatttattttactt 9540
 atccgtctac ttttttgcatttattttactt 9600
 gtcttttgcatttattttactt 9660
 ctacttaactt cccatggctt ttttttgcatttattttactt 9720
 atttgcaggcc ttttttgcatttattttactt 9780
 taattttttt ttttttgcatttattttactt 9840
 aggtggacta atattaaatg ttttttgcatttattttactt 9900
 tataacttaac ttttttgcatttattttactt 9960
 50 atacagaaaac ttttttgcatttattttactt 10020
 ttcttcacag ctttttgcatttattttactt 10080
 gaatataaaaa atgtatgtt ttttttgcatttattttactt 10140
 gcaacaagat atgtatgtt ttttttgcatttattttactt 10200
 atgtatgtt ttttttgcatttattttactt 10260
 55 gtttttgcatttattttactt 10320
 gtttttgcatttattttactt 10380
 caacctactt ttttttgcatttattttactt 10440
 acttgcgtt gtttttgcatttattttactt 10500
 aacaaggcag acaaccaaga ttttttgcatttattttactt 10560
 tacacaggaa ttttttgcatttattttactt 10620
 60 gggacatggaa ttttttgcatttattttactt 10680
 gatttgcgtt ttttttgcatttattttactt 10740
 atttcgagggt ttttttgcatttattttactt 10800
 gatcatgtc taggcctacc ttttttgcatttattttactt 10860
 aggtgtatgtc atatgtatgtt ttttttgcatttattttactt 10920
 65 tacttaatgt gaaagtttagg ttttttgcatttattttactt 10980
 aataagtcat ttttttgcatttattttactt 11040
 agactctctg accttagtaac ctttttgcatttattttactt 11040

tcaaggcattt cacccttgaag ttgagacgtt ttgttagtgt atactagttt atatgttgg 11100
 ggacatgttt atccagaaga tattcaggac tatttttgac tgggctaagg aattgattct 11160
 gatttagcact gttagtggc attgagttgc cittaggctt gaattggagt cacttgtata 11220
 tctcaaataa tgctggcctt tttaaaaag cccttgcctt ttatcaccct gtttctaca 11280
 5 taattttgt tcaaagaaat acttggttgg atctcctttt gacaacaata gcatgtttc 11340
 aagccatatt tttttcctt tttttttt ttttggttt ttcgagacag gtttctctg 11400
 tataggccctg gctgtccctgg aactcactt gtagaccagg ctggcctcga actcagaaat 11460
 10 ccgcctgcct ctgcctcctg agtgcggga ttaaaggcgt gcaccaccac gcctggctaa 11520
 gttggatatt ttgttatata actataacca atactaactc cactgggtgg atttttaatt 11580
 cagtcagtag tottaagtgg tctttatgg cccttcattt aatctactg ttcactctaa 11640
 cagaggctgt tggtactagt ggcacttaag caacttccta cgatataactt agcagattaa 11700
 gggtcaggga tagaaacttag tctagcggtt tgtataccta ccagctttat actaccttg 11760
 tctgatagaa atatttcagg acat 11784

15 <210> 107
 <211> 1458
 <212> DNA
 20 <213> Bacteriophage TP901-1
 <220>
 <221> CDS
 <222> (1)...(1455)

25 <400> 107
 atg act aag aaa gta gca atc tat aca cga gta tcc act act aac caa 48
 Met Thr Lys Lys Val Ala Ile Tyr Thr Arg Val Ser Thr Thr Asn Gln
 1 5 10 15

30 gca gag gaa ggg ttc tca att gat gag caa att gac cgt tta aca aaa 96
 Ala Glu Glu Gly Phe Ser Ile Asp Glu Gln Ile Asp Arg Leu Thr Lys
 20 25 30

35 tat gct gaa gca atg ggg tgg caa gta tct gat act tat act gat gct 144
 Tyr Ala Glu Ala Met Gly Trp Gln Val Ser Asp Thr Tyr Thr Asp Ala
 35 40 45

40 ggt ttt tca ggg gcc aaa ctt gaa cgc cca gca atg caa aga tta atc 192
 Gly Phe Ser Gly Ala Lys Leu Glu Arg Pro Ala Met Gln Arg Leu Ile
 50 55 60

aac gat atc gag aat aaa gct ttt gat aca gtt ctt gta tat aag cta 240
 Asn Asp Ile Glu Asn Lys Ala Phe Asp Thr Val Leu Val Tyr Lys Leu
 65 70 75 80

45 gac cgc ctt tca cgt agt gta aga gat act ctt tat ctt gtt aag gat 288
 Asp Arg Leu Ser Arg Ser Val Arg Asp Thr Leu Tyr Leu Val Lys Asp
 85 90 95

50 gtg ttc aca aaa aat aaa ata gac ttt atc tcg ctt aat gaa agt att 336
 Val Phe Thr Lys Asn Lys Ile Asp Phe Ile Ser Leu Asn Glu Ser Ile
 100 105 110

55 gat act tct tct gct atg ggt agc ttg ttt ctc act att ctt tct gca 384
 Asp Thr Ser Ser Ala Met Gly Ser Leu Phe Leu Thr Ile Leu Ser Ala
 115 120 125

60 att aat gag ttt gaa aga gag aat ata aaa gaa cgc atg act atg ggt 432
 Ile Asn Glu Phe Glu Arg Glu Asn Ile Lys Glu Arg Met Thr Met Gly
 130 135 140

aaa cta ggg cga gcg aaa tct ggt aag tct atg atg tgg act aag aca 480
 Lys Leu Gly Arg Ala Lys Ser Gly Lys Ser Met Met Trp Thr Lys Thr
 145 150 155 160

65 gct ttt ggg tat tac cac aac aga aag aca ggt ata tta gaa att gtt 528
 Ala Phe Gly Tyr Tyr His Asn Arg Lys Thr Gly Ile Leu Glu Ile Val

		92		
	165	170	175	
5	cct tta caa gct aca ata gtt gaa caa ata ttc act gat tat tta tca Pro Leu Gln Ala Thr Ile Val Glu Gln Ile Phe Thr Asp Tyr Leu Ser 180 185 190			576
10	gga ata tca ctt aca aaa tta aga gat aaa ctc aat gaa tct gga cac Gly Ile Ser Leu Thr Lys Leu Arg Asp Lys Leu Asn Glu Ser Gly His 195 200 205			624
15	atc ggt aaa gat ata ccg tgg tct tat cgt acc cta aga caa aca ctt Ile Gly Lys Asp Ile Pro Trp Ser Tyr Arg Thr Leu Arg Gln Thr Leu 210 215 220			672
20	gat aat cca gtt tac tgt ggt tat atc aaa ttt aag gac agc cta ttt Asp Asn Pro Val Tyr Cys Gly Tyr Ile Lys Phe Lys Asp Ser Leu Phe 225 230 235 240			720
25	gaa ggt atg cac aaa cca att atc cct tat gag act tat tta aaa gtt Glu Gly Met His Lys Pro Ile Ile Pro Tyr Glu Thr Tyr Leu Lys Val 245 250 255			768
30	caa aaa gag cta gaa gaa aga caa cag cag act tat gaa aga aat aac Gln Lys Glu Leu Glu Arg Gln Gln Thr Tyr Glu Arg Asn Asn 260 265 270			816
35	aac cct aga cct ttc caa gct aaa tat atg ctg tca ggg atg gca agg Asn Pro Arg Pro Phe Gln Ala Lys Tyr Met Leu Ser Gly Met Ala Arg 275 280 285			864
40	tgc ggt tac tgt gga gca cct tta aaa att gtt ctt ggc cac aaa aga Cys Gly Tyr Cys Gly Ala Pro Leu Lys Ile Val Leu Gly His Lys Arg 290 295 300			912
45	aaa gat gga agc cgc act atg aaa tat cac tgt gca aat aga ttt cct Lys Asp Gly Ser Arg Thr Met Lys Tyr His Cys Ala Asn Arg Phe Pro 305 310 315 320			960
50	cga aaa aca aaa gga att aca gta tat aat gac aat aaa aag tgt gat Arg Lys Thr Lys Gly Ile Thr Val Tyr Asn Asp Asn Lys Lys Cys Asp 325 330 335			1008
55	tca gga act tat gat tta agt aat tta gaa aat act gtt att gac aac Ser Gly Thr Tyr Asp Leu Ser Asn Leu Glu Asn Thr Val Ile Asp Asn 340 345 350			1056
60	ctg att gga ttt caa gaa aat aat gac tcc tta ttg aaa att atc aat Leu Ile Gly Phe Gln Glu Asn Asn Asp Ser Leu Leu Lys Ile Ile Asn 355 360 365			1104
65	ggc aac aac caa cct att ctt gat act tcg tca ttt aaa aag caa att Gly Asn Asn Gln Pro Ile Leu Asp Thr Ser Ser Phe Lys Lys Gln Ile 370 375 380			1152
	tca cag atc gat aaa aaa ata caa aag aac tct gat ttg tac cta aat Ser Gln Ile Asp Lys Lys Ile Gln Lys Asn Ser Asp Leu Tyr Leu Asn 385 390 395 400			1200
	gat ttt atc act atg gat gag ttg aaa gat cgt act gat tcc ctt cag Asp Phe Ile Thr Met Asp Glu Leu Lys Asp Arg Thr Asp Ser Leu Gln 405 410 415			1248
	gct gag aaa aag ctg ctt aaa gct aag att agc gaa aat aaa ttt aat Ala Glu Lys Lys Leu Leu Lys Ala Lys Ile Ser Glu Asn Lys Phe Asn 420 425 430			1296
	gac tct act gat gtt ttt gag tta gtt aaa act cag ttg ggc tca att			1344

Asp Ser Thr Asp Val Phe Glu Leu Val Lys Thr Gln Leu Gly Ser Ile			
435	440	445	
5 ccg att aat gaa cta tca tat gat aat aaa aag aaa atc gtc aac aac	1392		
Pro Ile Asn Glu Leu Ser Tyr Asp Asn Lys Lys Ile Val Asn Asn			
450	455	460	
10 ctt gta tca aag gtt gat gtt act gct gat aat gta gat atc ata ttt	1440		
Leu Val Ser Lys Val Asp Val Thr Ala Asp Asn Val Asp Ile Ile Phe			
465	470	475	
15 aaa ttc caa ctc gct taa	1458		
Lys Phe Gln Leu Ala			
485			
20 <210> 108			
<211> 485			
<212> PRT			
25 <213> Bacteriophage TP901-1			
<400> 108			
Met Thr Lys Lys Val Ala Ile Tyr Thr Arg Val Ser Thr Thr Asn Gln			
1	5	10	15
30 Ala Glu Glu Gly Phe Ser Ile Asp Glu Gln Ile Asp Arg Leu Thr Lys			
20	25	30	
35 Tyr Ala Glu Ala Met Gly Trp Gln Val Ser Asp Thr Tyr Thr Asp Ala			
35	40	45	
40 Gly Phe Ser Gly Ala Lys Leu Glu Arg Pro Ala Met Gln Arg Leu Ile			
50	55	60	
45 Asn Asp Ile Glu Asn Lys Ala Phe Asp Thr Val Leu Val Tyr Lys Leu			
65	70	75	80
50 Asp Arg Leu Ser Arg Ser Val Arg Asp Thr Leu Tyr Leu Val Lys Asp			
85	90	95	
40 Val Phe Thr Lys Asn Lys Ile Asp Phe Ile Ser Leu Asn Glu Ser Ile			
100	105	110	
45 Asp Thr Ser Ser Ala Met Gly Ser Leu Phe Leu Thr Ile Leu Ser Ala			
115	120	125	
50 Ile Asn Glu Phe Glu Arg Glu Asn Ile Lys Glu Arg Met Thr Met Gly			
130	135	140	
55 Lys Leu Gly Arg Ala Lys Ser Gly Lys Ser Met Met Trp Thr Lys Thr			
145	150	155	160
55 Ala Phe Gly Tyr Tyr His Asn Arg Lys Thr Gly Ile Leu Glu Ile Val			
165	170	175	
60 Pro Leu Gln Ala Thr Ile Val Glu Gln Ile Phe Thr Asp Tyr Leu Ser			
180	185	190	
60 Gly Ile Ser Leu Thr Lys Leu Arg Asp Lys Leu Asn Glu Ser Gly His			
195	200	205	
65 Ile Gly Lys Asp Ile Pro Trp Ser Tyr Arg Thr Leu Arg Gln Thr Leu			
210	215	220	
65 Asp Asn Pro Val Tyr Cys Gly Tyr Ile Lys Phe Lys Asp Ser Leu Phe			
225	230	235	240

94

Glu Gly Met His Lys Pro Ile Ile Pro Tyr Glu Thr Tyr Leu Lys Val
 245 250 255

5 Gln Lys Glu Leu Glu Glu Arg Gln Gln Gln Thr Tyr Glu Arg Asn Asn
 260 265 270

Asn Pro Arg Pro Phe Gln Ala Lys Tyr Met Leu Ser Gly Met Ala Arg
 275 280 285

10 Cys Gly Tyr Cys Gly Ala Pro Leu Lys Ile Val Leu Gly His Lys Arg
 290 295 300

Lys Asp Gly Ser Arg Thr Met Lys Tyr His Cys Ala Asn Arg Phe Pro
 305 310 315 320

15 Arg Lys Thr Lys Gly Ile Thr Val Tyr Asn Asp Asn Lys Lys Cys Asp
 325 330 335

20 Ser Gly Thr Tyr Asp Leu Ser Asn Leu Glu Asn Thr Val Ile Asp Asn
 340 345 350

Leu Ile Gly Phe Gln Glu Asn Asn Asp Ser Leu Leu Lys Ile Ile Asn
 355 360 365

25 Gly Asn Asn Gln Pro Ile Leu Asp Thr Ser Ser Phe Lys Lys Gln Ile
 370 375 380

30 Ser Gln Ile Asp Lys Lys Ile Gln Lys Asn Ser Asp Leu Tyr Leu Asn
 385 390 395 400

Asp Phe Ile Thr Met Asp Glu Leu Lys Asp Arg Thr Asp Ser Leu Gln
 405 410 415

35 Ala Glu Lys Lys Leu Leu Lys Ala Lys Ile Ser Glu Asn Lys Phe Asn
 420 425 430

Asp Ser Thr Asp Val Phe Glu Leu Val Lys Thr Gln Leu Gly Ser Ile
 435 440 445

40 Pro Ile Asn Glu Leu Ser Tyr Asp Asn Lys Lys Ile Val Asn Asn
 450 455 460

Leu Val Ser Lys Val Asp Val Thr Ala Asp Asn Val Asp Ile Ile Phe
 465 470 475 480

45 Lys Phe Gln Leu Ala
 485

Fig. 1

1/9

A. pCMV-C31Int(wt):



B. pCMV-C31Int(N-NLS):



C. pCMV-C31Int(C-NLS):



D. pCMV-Cre:



E. pRK64:

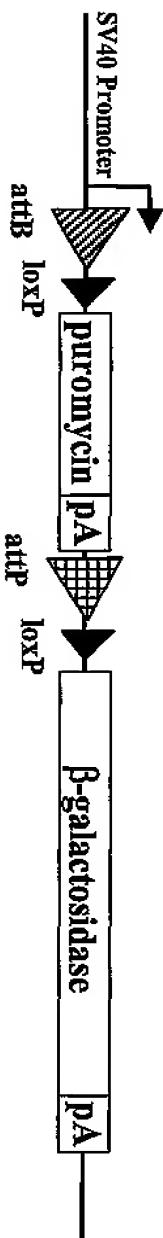


Fig. 2

Sample	RLU (β Gal)	RLU (Luciferase)	RLU x10 ⁵ (Gal/Luc)	RLU x10 ⁵ (Gal/Luc)	relative activity
1) pUC19 only	433 \pm 37	1818700 \pm 328970	24 \pm 4	10 ⁴	5 \times 10 ⁴
2) pRK64(Δ Cre)	784932 \pm 290524	1657062 \pm 526562	46975 \pm 3696		1
3) pRK64 only	1936 \pm 262	988144 \pm 175116	204 \pm 62		0.004
4) pCMV-C31Int(wt) 0.5 ng	73318 \pm 19084	677861 \pm 145341	10774 \pm 972		0.23
5) pCMV-C31Int(wt) 1 ng	75838 \pm 12628	527237 \pm 53846	14412 \pm 2050		0.3
6) pCMV-C31Int(NNLS) 0.5 ng	158402 \pm 75870	2560450 \pm 736186	15104 \pm 3041		0.32
7) pCMV-C31Int(NNLS) 1 ng	206857 \pm 76733	2677621 \pm 504285	17029 \pm 2246		0.36
8) pCMV-C31Int(CNLS) 0.5 ng	274192 \pm 78937	1173932 \pm 291315	23299 \pm 3194		0.5
9) pCMV-C31Int(CNLS) 1 ng	262169 \pm 60583	864752 \pm 229935	30560 \pm 1585		0.65
10) pCMV-Cre(NNLS) 0.5 ng	231200 \pm 96741	763121 \pm 280687	29595 \pm 4632		0.63
11) pCMV-Cre(NNLS) 1 ng	297760 \pm 83363	868905 \pm 196404	33872 \pm 2609		0.72

Fig. 3

	RLU (β Gal)	RLU (Luciferase)	RLU $\times 10^5$ (Gal/Luci)
1) pPGKnefD (reporter) only	1324 \pm 876	3631598 \pm 903012	34 \pm 18
2) pCMV-XisA 25 ng	4650 \pm 2273	2741969 \pm 667568	164 \pm 54
3) pCMV-XisA 100 ng	17529 \pm 9304	3798872 \pm 1238020	443 \pm 151
4) pCMV-XisA(NNLS) 25 ng	4060 \pm 1376	2471695 \pm 611351	163 \pm 36
5) pCMV-XisA(NNLS) 100 ng	17801 \pm 3892	3570103 \pm 750628	500 \pm 65
6) pPGKattA (reporter) only	754 \pm 70	195822 \pm 81858	755 \pm 601
7) pCMV-SSV 10 ng	925 \pm 273	119043 \pm 67451	906 \pm 316
8) pCMV-SSV 20 ng	1033 \pm 270	122557 \pm 30054	879 \pm 291
9) pCMV-SSV(NNLS) 10 ng	1108 \pm 367	174380 \pm 58876	694 \pm 345
10) pCMV-SSV(NNLS) 20 ng	1306 \pm 383	211182 \pm 101011	874 \pm 741

4/9

Fig. 4

Sample	RLU (β Gal)	RLU (β Gal)	relative activity
1) pUC19 only	754 \pm 44	5 \times 10 ³	
2) pCMV-C31Int(wt)	32 ng	1386 \pm 174	0.04
3) pCMV-C31Int(wt)	64 ng	3783 \pm 1537	0.1
4) pCMV-C31Int(NNLS)	32 ng	7125 \pm 1474	0.19
5) pCMV-C31Int(NNLS)	64 ng	8206 \pm 2210	0.22
6) pCMV-C31Int(CNLS)	32 ng	17624 \pm 5578	0.48
7) pCMV-C31Int(CNLS)	64 ng	28849 \pm 6623	0.78
8) pCMV-Cre(NNLS)	32 ng	27064 \pm 3769	1
9) pCMV-Cre(NNLS)	64 ng	36823 \pm 3993	

A.: Nontransfected control

B.: pCMV-Cre

C.: pCMV-C31Int(NLS)

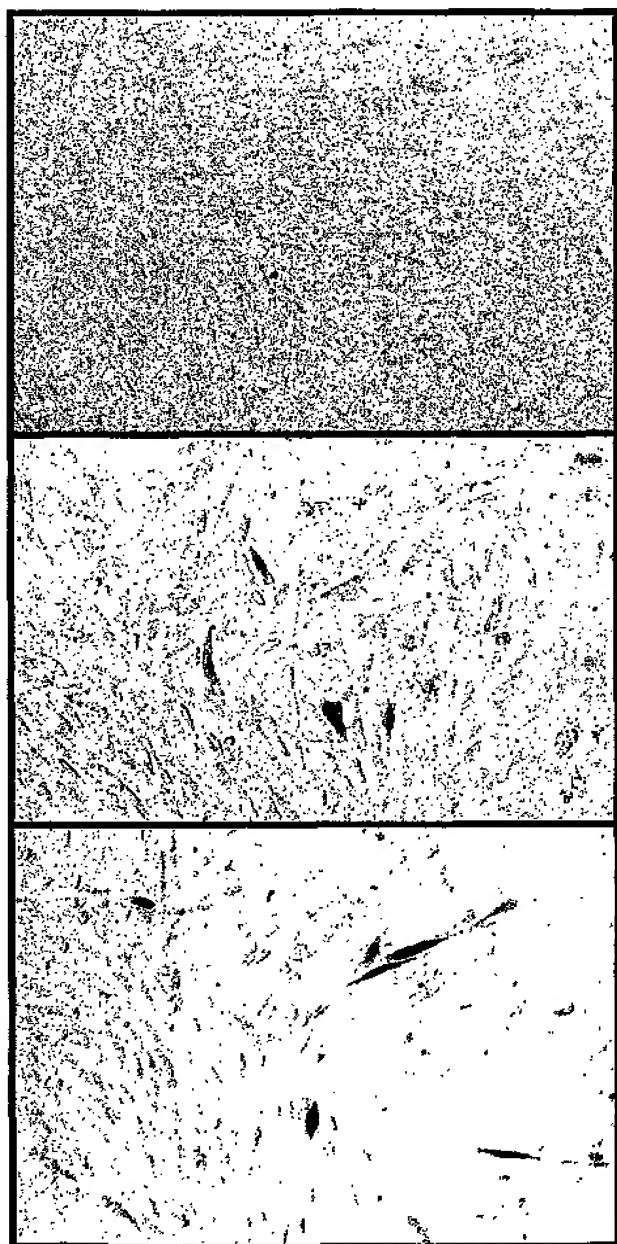


Fig. 5

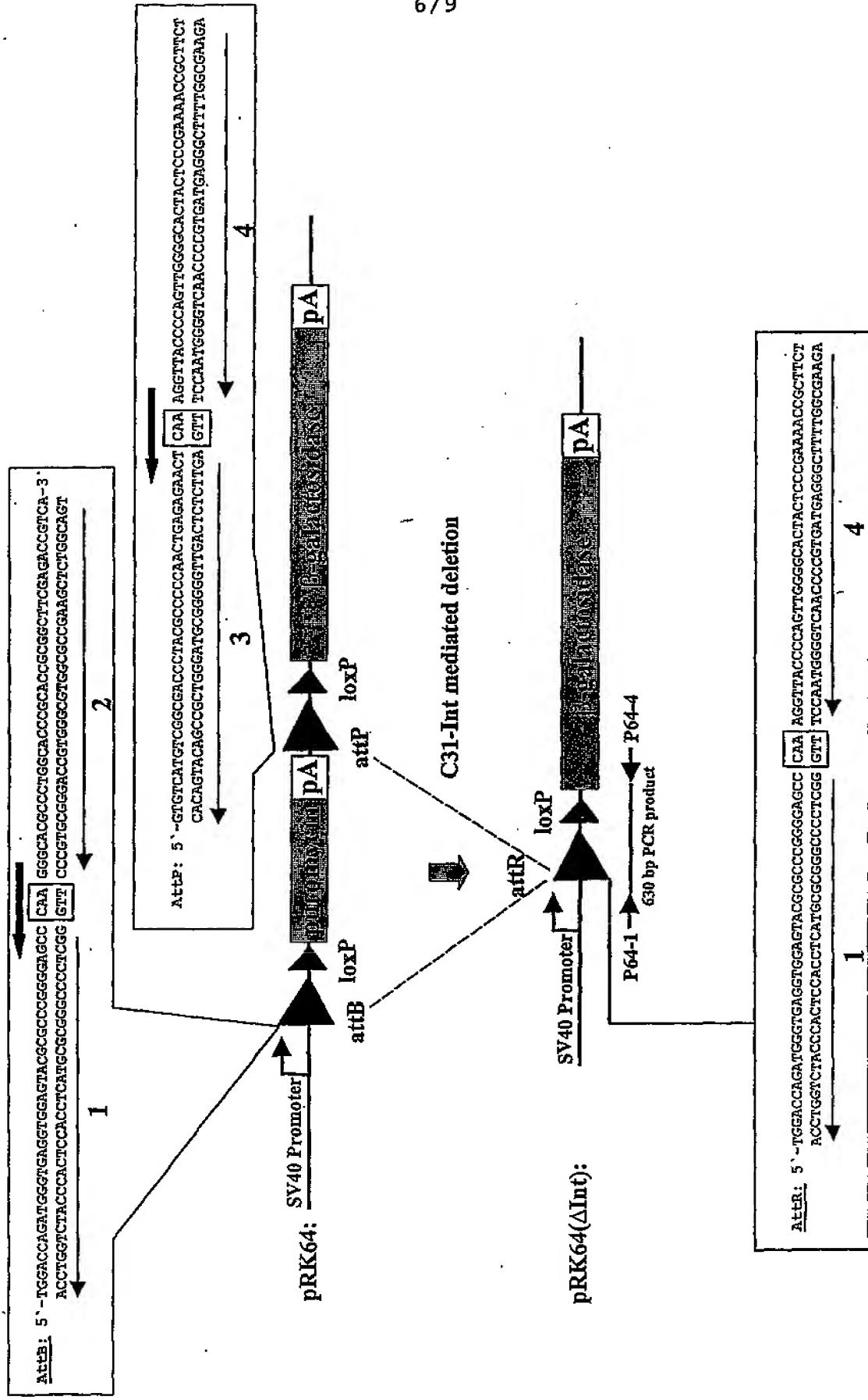


Fig. 6

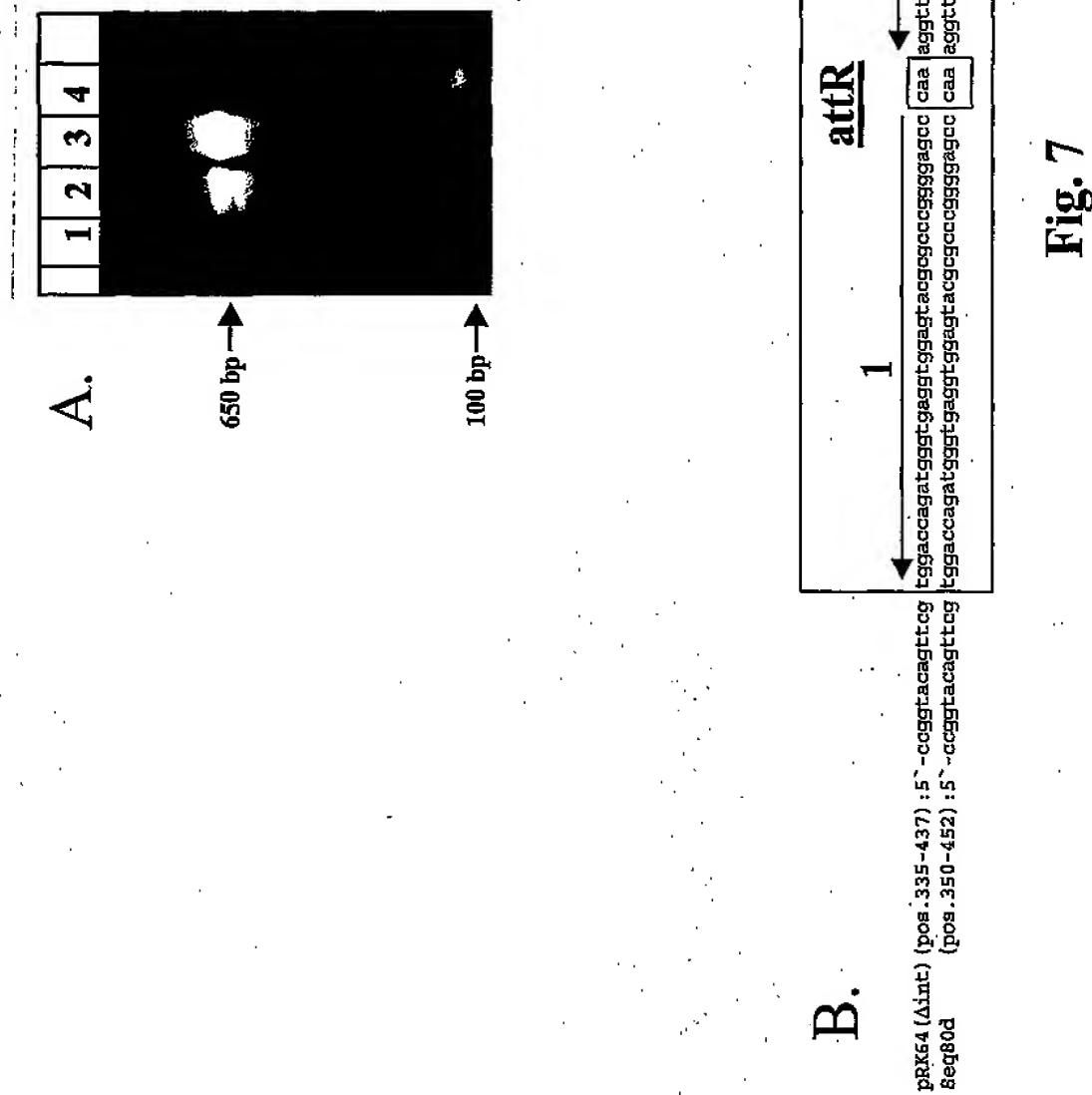
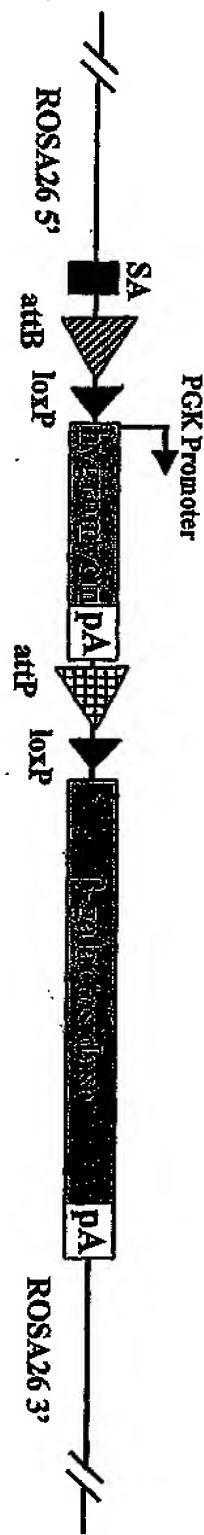


Fig. 7

Fig. 8



9 / 9

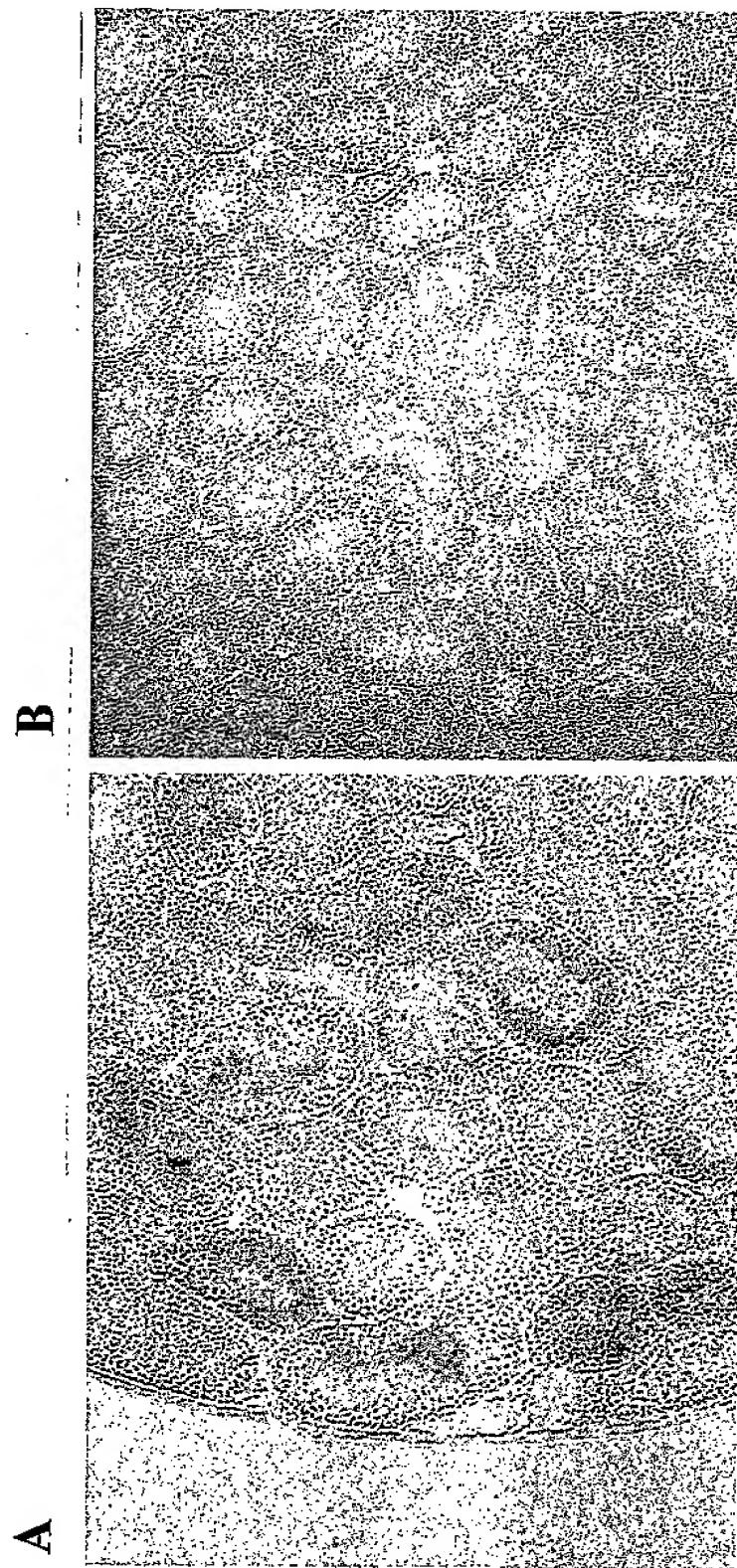


Fig. 9